COGNTIVE 2021

The Thirteenth International Conference on Advanced Cognitive Technologies and Applications

ISBN: 978-1-61208-847-1

April 18 - 22, 2021

COGNITIVE 2021 Editors

Olga Chernavskaya, P.N. Lebedev Physical Institute, Moscow, Russia
Joao E. Kogler Jr, University of Sao Paulo, Brazil
Igor Val Danilov, Academic Center for Coherent Intelligence, Latvia
Oana Dini, IARIA, EU/USA

COGNITIVE 2021

Forward

The Thirteenth International Conference on Advanced Cognitive Technologies and Applications (COGNITIVE 2021), held on April 18 - 22, 2021, targeted advanced concepts, solutions and applications of artificial intelligence, knowledge processing, agents, as key-players, and autonomy as manifestation of self-organized entities and systems. The advances in applying ontology and semantics concepts, web-oriented agents, ambient intelligence, and coordination between autonomous entities led to different solutions on knowledge discovery, learning, and social solutions.

The conference had the following tracks:

- Brain information processing and informatics
- Artificial intelligence and cognition
- Agent-based adaptive systems
- Applications

Similar to the previous edition, this event attracted excellent contributions and active participation from all over the world. We were very pleased to receive top quality contributions.

We take here the opportunity to warmly thank all the members of the COGNITIVE 2021 technical program committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and effort to contribute to COGNITIVE 2021. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations and sponsors. We also gratefully thank the members of the COGNITIVE 2021 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope COGNITIVE 2021 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the area of cognitive technologies and applications.

COGNITIVE 2021 General Chair

Jaime Lloret Mauri, Universitat Politecnica de Valencia, Spain

COGNITIVE 2021 Steering Committee

Charlotte Sennersten, CSIRO Mineral Resources, Australia
Jayfus Tucker Doswell, The Juxtopia Group, Inc., USA

COGNITIVE 2021 Publicity Chair

Jose Luis García, Universitat Politecnica de Valencia, Spain
Lorena Parra, Universitat Politecnica de Valencia, Spain

COGNITIVE 2021 Industry/Research Advisory Committee

Tiago Thompsen Primo, Samsung Research Institute, Brazil
Kazuhisa Miwa, Nagoya University, Japan
Olga Chernavskaya, Lebeved Physical Institute, Moscow, Russia
Paul Smart, University of Southampton, UK

COGNITIVE 2021

Committee

COGNITIVE 2021 General Chair

Jaime Lloret Mauri, Universitat Politecnica de Valencia, Spain

COGNITIVE 2021 Steering Committee

Charlotte Sennersten, CSIRO Mineral Resources, Australia
Jayfus Tucker Doswell, The Juxtopia Group, Inc., USA

COGNITIVE 2021 Industry/Research Advisory Committee

Tiago Thompsen Primo, Samsung Research Institute, Brazil
Kazuhisa Miwa, Nagoya University, Japan
Olga Chernavskaya, Lebeved Physical Institute, Moscow, Russia
Paul Smart, University of Southampton, UK

COGNITIVE 2021 Publicity Chairs

Jose Luis García, Universitat Politecnica de Valencia, Spain
Lorena Parra, Universitat Politecnica de Valencia, Spain

COGNITIVE 2021 Technical Program Committee

Witold Abramowicz, University of Economics and Business, Poland
Thomas Agotnes, University of Bergen, Norway
Vered Aharonson, University of the Witwatersrand, Johannesburg, South Africa
Sadam Al-Azani, King Fahd University of Petroleum & Minerals, Saudi Arabia
Piotr Artiemjew, University of Warmia and Masuria in Olsztyn, Poland
Petr Berka, University of Economics, Prague, Czech Republic
Ateet Bhalla, Independent Consultant, India
Mahdi Bidar, University of Regina, Canada
Guy Andre Boy, CentraleSupélec, LGI,Paris Saclay University /ESTIA Institute of Technology, France
Dilyana Budakova, Technical University of Sofia - Branch Plovdiv, Bulgaria
Valerie Camps, Paul Sabatier University - IRIT, Toulouse, France
Yaser Chaaban, Leibniz University of Hanover, Germany
Olga Chernavskaya, P. N. Lebeved Physical Institute, Moscow, Russia
Helder Coelho, Universidade de Lisboa, Portugal
Igor Val Danilov, Academic Center for Coherent Intelligence, Latvia
Angel P. del Pobil, Jaume I University, Spain
Soumyabrata Dev, University College Dublin, Ireland

Razieh Saremi, Stevens Institute of Technology, USA
Charlotte Sennersten, CSIRO Mineral Resources, Australia
Ljiljana Šerić, University of Split, Croatia
Paul Smart, University of Southampton, UK
Priyanka Srivastava, International Institute of Information Technology, Hyderabad (IIIT-H), India
Stanimir Stoyanov, Plovdiv University "Paisii Hilendarski", Bulgaria
Nasseh Tabrizi, East Carolina University, USA
Tiago Thompsen Primo, Samsung Research Institute, Brazil
Gary Ushaw, Newcastle University, UK
Jaap van den Herik, Leiden Centre of Data Science (LCDS) | Leiden University, Leiden, The Netherlands
S.Vidhusha, SSN College of Engineering, Chennai, India
Emilio Vivancos, Valencian Research Institute for Artificial Intelligence (VRAIN) | Universitat Politecnica de Valencia, Spain
Xianzhi Wang, University of Technology Sydney, Australia
Ye Yang, Stevens Institute of Technology, USA
Sule Yildirim Yayilgan, NTNU, Norway
Besma Zeddini, EISTI, France

Copyright Information

Table of Contents

Intentionality vs Chaos

Brain Connectivity through Emotions and Cooperation Levels beyond Sensory Modalities.

Igor Val Danilov
Academic Center for Coherent Intelligence
ACCI
Roma, Italy
e-mail: igor_val.danilov@acci.center

Sandra Mihailova
Rīga Stradiņš University
RSU
Riga, Latvia
e-mail: sandra.mihailova@rsu.lv

***Abstract*—Empirical evidence shows the efficiency of coordinated interaction in mother-infant dyads through unintentional movements: social entrainment, early imitation. The growing body of the literature evidently shows an impact of arousal on group performance and spreading emotion from one individual to another organism, called emotional contagion. The emotion sharing somehow stimulates shared intentionality in individuals of "primary group". Although there is increasing evidence of consistency between some "motion" and "emotion" concepts, research demonstrating the synergy of the integrative process of all these five concepts is limited. This study presents a dynamic integrated model of motion-emotions synergy that combines and manages all the above-mentioned appearances–the Model of Coherent Intelligence. This model is supported by arguments that are strictly based on experimental evidence in the literature and 24 online experiments in 2020. Primary groups show empirical evidence of a more significant accuracy in problem-solving in the coherent intelligence state. In specific, we conducted 13 online experiments in dyads (116 subjects) with *P*-value < 0.001, and 7 experiments in primary group adults (41 subjects) with the *P*-value < 0.002. Experiments with not well-known adults from the secondary group show the effect only within task UL3 (41 subjects in experiment No.12 with translation of unfamiliar language). Non-semantic tasks–SL3 (synthetic language) and US3 (two-color symbols)–did not stimulate the effect in 3 experiments with 207 secondary group subjects (unfamiliar students). The outcome demonstrates inter-brain connectivity through ongoing emotions and motion dynamics, creating cooperation levels beyond sensory modalities.**

Keywords-social cognition; coherent intelligence; embodied cognition; emotion contagion; imitation; interactional synchrony; social entrainment.

I. Introduction

Empirical evidence shows the efficiency of coordinated interaction in mother-infant dyads through unintentional movements: social entrainment [1][2], early imitation [3][4], and interactional synchrony [5][6]. The growing body of the literature evidently shows an impact of arousal on group performance [7]-[9] and spreading emotion from one individual to another organism [10]-[12], called emotional contagion. That is, emotion sharing somehow stimulates sharing intentionality in individuals of "primary group"[13]. Although there is increasing evidence of consistency between some "motion" and "emotion" concepts, research demonstrating the synergy of the integrative process of all these five concepts is limited.

Recent research presented a more significant accuracy level when participants independently completed similar tasks parallel with confederates who were primed with the correct answer [14]. This research design stimulated their emotional arousal and interactional synchrony in face-to-face performances. The current paper presents the outcome of 24 online experiments in 2020, which were designed following the Model of Coherent Intelligence (MCI).

Section II presents the hypothesis of how ongoing social dynamics can create a coherent mental process in groups. This Model of Coherent Intelligence (MCI) argues that social interaction shapes organisms' intentionality. Section III contains research data of 24 online experiments, their research problem, paradigm, and procedure. Section IV discusses limitations and difficulties of this research. Section V elaborates all findings, describing their meanings.

II. The Model of Coherent Intelligence

According to Danilov and Mihailova [15], the MCI assumes that ongoing social dynamics create a coherent mental process in dyads (primary group) where movement coordination is cyclically enhanced under ever-growing arousal. A supranormal environment, e.g., first hours after birth, stimulates supranormal sensation in dyads. This can push the inherited mechanism of social entrainment of infants to the rhythm of the mother. Both the supranormal sensation and social entrainment may stimulate the common emotional arousal. The latter is increased by the ongoing supranormal sensation and the occurring rhythm of arbitrary movements of the infant. The continuing supranormal sensation and ever-increasing arousal of the infant and the mother, along with the rhythm of the infant's unintentional movements, stimulate early imitation and emotional contagion. How the infant captures and reproduces the kinematic of movements. The MCI proposes that common emotional arousal together with the identical rhythm create coherent mental processes in dyads–Coherent Intelligence–as it is shown in Figure 1.

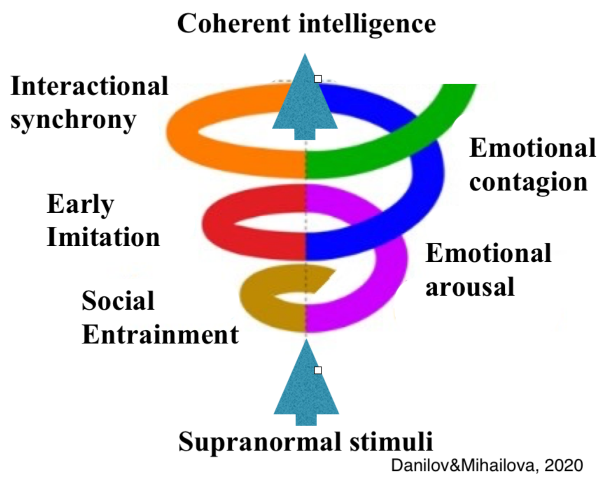


Figure 1. The model of coherent intelligence

At Stage 3 of the Model of Hierarchical Complexity MHC [16], organisms (such as human infants) do not maintain bilateral communication. According to Danilov and Mihailova [17], individuals are able to interact by distinguishing perceptual signals of identical modality by their significance. This is the foundation of intentionality. This ability can contribute to ostensive cues, categorizing reality. After all, this meaningless interaction modifies into communication when individuals imbue perceptual impulses with mutually implied meanings, cascading their signals in response to the history of relations between them. In such a manner, organisms at Stage 3 of the MHC development begin to categorize the chaos of sensory stimuli. The hypothesis of the MCI argues that social interaction shapes organisms' intentionality, promoting similar categorization of stimuli in intimately related individuals with shared social routine and interests.

III. EXPERIMENTS

In 2020, we conducted 24 experiments with 407 subjects to test the MCI hypothesis of whether this effect also appears online.

A. The Research Problem

The research problem of these 24 online experiments was whether or not unprimed participants show a more significant accuracy level when they complete the thought task simultaneously with confederates who are primed with the correct answer; if they were emotionally stimulated and completed the tasks without communication. Would a confident knowledge of the confederates on the tasks help target participants to solve these unintelligible problems without communication, when they simultaneously pass the same testing?

B. Experiment Paradigm and the Procedure

The within subjects paradigm was applied to verify the difference between the correct responses of the participants (unprimed subjects) to the "primed block" and "unprimed block" of tasks. Specifically, subjects (unprimed participants and primed confederates) with their computers joined the experiment through an online conference. They were also asked to prepare their mobile phones to complete the experimental tasks during the online conference by connecting to the experiment's website via their mobile phones. As soon as they all are online, experimenters divided subjects into two groups (participants and confederates) and informed them about the task.

Then, all subjects were asked to use their mobile phones to contact the experiment's website for completing the test. As soon as all subjects were connected via their mobiles to the website, testing began simultaneously. The experiment design stimulated arousal by the experiment's unusual situation and rhythmically changing red/purple colors of the mobiles' screens. Many studies on emotion arousal in learning show increased participants' cortisol levels during experiments [15]. It is highly likely that unintelligible tasks also contributed to emotional excitement in subjects as well as the participation of strangers (experimenters) in these performances. Interactional synchrony in subjects was stimulated by rhythmically changing colors of the mobiles' screen (80 bpm), that was identical for all.

During testing, the website simultaneously presented to all subjects 10 tasks. The all tasks design was the same for all subjects (participants and confederates)–all subjects saw the same picture with the similar mapping of the task and its answer options. The design of each task promoted the same geometrical navigation on the screen for all. The mapping of Unfamiliar Language (UL) task presented 10 answer options in one line on the screen of the mobile phones. The mapping of unfamiliar Synthetic Language (SL) task presented 8 answer options located on the square's perimeter on the screen. The mapping of Unintelligible Symbols (US) task presented on the screens 4 answer options located on the square's corners. That is, each task had the same for all subjects task-options mapping design, and answer options were in the same place on the mobile phone screens for all subjects (participants and confederates). In such a manner, we designed the same geometrical point of the correct answer on the screens for all participants and confederates.

The confederates were asked to follow hints on the right answers, solving tasks. They received hints on all even-numbered tasks (2-4-6-8-10)–"primed block" of tasks. The unprimed tasks for the confederates were all un-even (odd-numbered) tasks (1-3-5-7-9)–"unprimed block" of tasks. They did not receive instructions on odd-numbered tasks. In evaluating the outcome, the "unprimed block" of uneven tasks (1-3-5-7-9) became the baseline of the experiment. Unprimed participants did not receive any instruction on any task; both the unprimed block (1-3-5-7-9) and the primed block (2-4-6-8-10) were unintelligible tasks for them. That is, we tested whether or not unprimed participants (target participant) would be more accurate in solving unintelligible tasks when they were doing this simultaneously with confederates who knew the correct answer. Would unprimed subjects' (participants') results be better in the primed block than in the unprimed one? Would a confident knowledge of confederates on primed block tasks help unprimed subjects (target participants) to solve them without communication?

The result was estimated by two values: (1) Rb – the ratio between the correct responses of the unprimed participants to the "primed block" and "unprimed block" of tasks, (2) Rch – the ratio between the correct responses of the unprimed participants to the "primed block" and possible responses by chance.

B. *Experiments with Translation of Unfamiliar Language (UL)*

UL1: We conducted 6 online experiments with 22 dyads (44 person); we tested subjects who are typical representatives of the primary group. The task was to choose the correct translation of an unfamiliar language from 10 variants from the list of 10 options. Before the experiment, we asked the subjects if they spoke the investigated language to make sure that this language was unfamiliar for them. Each dyad was divided into unprimed participants (a child) and primed confederates (her mother). They were asked to avoid any communication between them during testing. Dyads were required to translate unfamiliar foreign words themselves (independently) by choosing one correct translation from 10 variants in their native language in a congruent design and, with the opposite task, in an incongruent one. The confederates received correct answers on "primed block" of tasks (all even tasks 2-4-6-8-10). These online experiments in different languages found evidence of an increase of Rb=48% in a group performance between "primed block" of tasks and "unprimed block", and an increase of Rch=90% above chance in "primed block", the results are in Table 1.

UL2: The same tasks under the similar procedure were tested with 24 adults: students and groups of friends. These subjects cannot be attributed as typical representatives of the primary group; nevertheless, many of them stay in a coordinated state of social entrainment with their fellow, as we believe. We estimated them as primary group. The subjects of five experiments No. from 7 (18/04/2020) to 11 (12/05/2020) had been studying together since many years. These online experiments with different unfamiliar languages found evidence of an increase of Rb=143% in a group performance between "primed block" of tasks and "unprimed block", and an increase of Rch=216% above chance in "primed block", the results are in Table 1.

UL3: We repeated the same procedure in experiment No. 12 (12/05/2020) with subjects who were 41 second-year university students. They knew each other not more then two years and it seemed difficult to define their biological state as close to the social entrainment. We estimated them as secondary group. The results of 4 students were excluded from the outcome, since they speak the examined language. The online experiment No.12 shows evidence of an increase of Rb=133% in a group performance between "primed block" of tasks and "unprimed block", and an increase of Rch=380% above chance in "primed block", the results are in Table 1.

C. *Experiments with a Rebus from Synthetic Language (SL)*

SL1: We conducted 4 online experiments with 23 children and 19 mothers (specifically 19 families). The task was to choose the correct version of the rebus from 8 options placed around the square's perimeter. The rebus consisted of unknown symbols from synthetic language created especially for this experiment. These online experiments found evidence of an increase of Rb=394% in a group performance between "primed block" of tasks and "unprimed block", and an increase of Rch=42% above chance in "primed block", the results are in Table 1.

SL2: We tested the same tasks with 7 adults (M=18): students of the last year of high school. We believe they were in a coordinated state of social entrainment because of many years studying together under the same schedule. We estimated them as primary group. These online experiments found evidence of an increase of Rb=300% in a group performance between "primed block" of tasks and "unprimed block", and an increase of Rch=28% above chance in "primed block", the results are in Table 1.

SL3: We tested the same tasks with 56 adults (M=21): students of the second year of university. We estimated them as secondary group. They knew each other for not more than two years, and it does not seem easy to define their biological state as close to social entrainment. These online experiments did not find evidence of Rb's increase in a group performance between "primed block" of tasks and "unprimed block". At the same time participants' outcome in both conditions was Rch=28% above chance, the results are in Table 1.

D. *Experiments with Unintelligible Symbols US*

US1: We conducted 3 online experiments with 17 children (M=9) and 13 mothers (M=40); specifically, there were 13 families. Their task was to choose the two-color symbol related to one of the natural numbers from 1 to 5 among four options. In each task, four answer options (different two-color symbols) were located in the corners of the square's perimeter. These five two-color symbols related to a natural number from 1 to 5 were created specifically for this experiment. These symbols consisted of two colored circles (one in the other). The meanings of these unfamiliar symbols were unintelligible for subjects. Experimenters asked subjects to solve the tasks applying different strategies. Participants (unprimed subjects) were asked to guess the correct answer intuitively. In contrast, confederates were asked to solve the problems rationally, following hints on the correct answer. These online experiments in different languages found evidence of an increase of Rb=123% in a group performance between "primed block" of tasks and "unprimed block", and an increase of Rch=32% above chance in "primed block", the results are in Table 1.

US2: We conducted 1 online experiment with a group of 10 friend adults: 4 confederates and 6 participants (M=30). Their task was the same as of US1: to choose the two-color symbol related to one of the natural numbers from 1 to 5 among four options. Their meanings were unintelligible for subjects. Experimenters asked subjects to solve the tasks applying different strategies. Participants (unprimed subjects) were asked to guess the correct answer intuitively. In contrast, confederates were asked to solve the problems rationally, following hints on the correct answer. These online experiments in different languages found evidence of an increase of Rb=127% in a group performance between "primed block" of tasks and "unprimed block", and an increase of Rch=30% above chance in "primed block", the results are in Table 1.

 ISBN: 978-1-61208-847-1

US3: We conducted 2 online experiments with 151 students from the first year of the university (M=19). We estimated them as secondary group. They knew each other for not more than a few days (if they knew each other, they did not meet because of the online university course), and it seems impossible to define their biological state as close to the social entrainment. Their task was the same as of US1: to choose the two-color symbol related to one of the natural numbers from 1 to 5 among four options. Experimenters asked subjects again to solve the tasks applying different strategies. Participants (unprimed subjects) were asked to guess the correct answer intuitively.

In contrast, confederates were asked to solve the problems rationally, following hints on the correct answer. These online experiments in different languages found evidence of an increase of Rb=3% in a group performance between "primed block" of tasks and "unprimed block". Their results Rch were below chance, see Table 1.

E. Results

The 20 experiments in subjects from the primary group includes 13 experiments in dyads (with 58 mothers and 68 children), and 7 experiments with 41 adults. The 4 experiments in subjects from the secondary group with 250 adults showed the effect only in UL3 task (a translation of an unfamiliar language). Other experiments in secondary group with the tasks SL3 and US3 did not show the effect.

The results are presented in the Table 1. There are several abbreviations to note: UL – the experiments with translation of an unfamiliar language; SL – the experiments with a rebus from unknown symbols of a synthetic language; US – the experiments with two-color round symbols; Rb, equation (1) – the ratio between the correct responses of the unprimed participants to the "primed block" and "unprimed block" of tasks; Rch, equation (2) – the ratio between the correct responses of the unprimed participants to the "primed block" and possible responses by chance; Mp – mean primed; Mb – mean baseline (unprimed); E – estimated by chance; O – observed results (both Mp and Mb); *P*-value – the significance of results, rejecting the null hypothesis; and values of the Chi-squared distribution is χ^2 .

IV. Discussion

The limitations of the unfamiliar language task (UL) and synthetic language task (SL) are grounded in the life experience of subjects. Translation of an unfamiliar language (or solving a rebus) is challenging; however, subjects could casually hear some foreign words in the past and/or create associative relationships with words/symbols they already knew.

$$100\% \times (Mp - Mb)/Mb = Rb. \quad (1)$$

$$100\% \times (Mp - E)/E = Rch. \quad (2)$$

$$\Sigma(O - E)^2/E = \chi^2. \quad (3)$$

Table 1. Results of 24 experiments in 2020

Group	The Ratio of Correct Responses					
	Ratio	***Task UL, %***	***Task SL, %***	***Task US, %***	χ^2	***P-value***
1. Dyads, 116 subjects	Rb	48	394	123	16.142	< 0.001
	Rch	90	42	32		
2. Primary group, 41 subjects	Rb	143	300	127	13.493	< 0.002
	Rch	216	28	20		
3 Secondary group, 250 subjects	Rb	–	–8	3	O.083	< 0.975
	Rch	–	31	–9		
	Rb	133	–	–	250.624	< 0.001
	Rch	380	–	–		

This implicit knowledge cannot be completely excluded from the outcome. Even though we selected unfamiliar foreign words (as it seemed to us) verifying this linguistic task with a control group, this did not exclude such cultural influence on subjects' results. Possible past experience and/or associative relationships between words/symbols could make an adjustment to the results. Nevertheless, it seems uncontroversial to say that young children acquire knowledge through a communication environment: language and other communicative signs. The current paper explores the modalities of social interaction that help organisms to assimilate knowledge. Therefore we propose to take into account this outcome because language is a typical communicative cue for children. It is possible to suppose that a communicative environment–symbols' domain–can enhance non-perceptual interaction. Principally we suggest mentioning this since the control group did not show any difference in results between even-numbered tasks and uneven (odd-numbered) tasks, testing all of them under unprimed conditions. This control group outcome may mean that the set of foreign words was unfamiliar for the particular subjects of the control group and may provide hope that this was a spread case also for other subjects. In contrast, the tasks with two-color round symbols (US) could create less association with previous knowledge in participants. From this perspective, the task US could show pure non-perceptual social interaction. These limitations can describe the difference in results between UL, SL, and US tasks in different groups. For instance, the primary group (dyads and adults) perform better linguistic task UL than the tasks with rebus from unknown symbols SL and two-colored symbols US, showing in the UL better results above chance.

It seems uncontroversial to say that the third task US with two-color round symbols excluded participants' experience (previous knowledge) from problem-solving to a greater extent. Comparing results between different tasks and groups shows the lowest increase of the Rch coefficient – the ratio between the correct responses between the "primed block" and responses by chance.

 ISBN: 978-1-61208-847-1

One of the research difficulties was to ensure the intentionality of the primed subjects (confederates), since the only following the instructions on the correct answers was not too exciting for them. We expected their mental collaboration instead of indifferent action in choosing of correct option. Therefore, for each experiment we created the special game for confederates depending on their personal interests. Although, in the sense of a person's unexpected choice, none can be sure on what to expect from a person, within reasonable limits, of course. Human uncertainty creates the problem for any research in psychology. Frankness, sincerity and involvement in the experiment are most influent factors of the testee's impact on the results of research. The experimenter can never exactly know the real intention and involvement of the examinee. We believe that the difference in the outcomes of the different groups also shows the participants involvement in the process.

The outcome was questioned: whether the correlation of results between teams of subjects (participants and confederates) is an evident pattern, or it happened by chance. The hypothesis evaluation using the *P*-value shows the significance of results in experiments with dyads and adults attributed to the primary group. The 20 experiments with subjects from primary group show: (i) the *P*-value < 0,001 in 13 experiments in dyads; and (ii) the *P*-value < 0,002 in 7 experiments in adults. We believe this outcome is statistically significant, rejecting the null hypothesis in subjects attributed to the primary group.

V. Conclusions

We conducted 24 online experiments with subjects of the primary group (20 experiments, 157 subjects) and subjects of the secondary group (4 experiments, 250 subjects). To our mind, the results of these online experiments support a hypothesis of inter-brain connectivity, which appears in individuals of primary groups (including dyads) at the beginning of cognition and lasts the entire social life. The unprimed subjects (participants) attributed to the primary group showed a more significant level of accuracy when they completed a thought task in the presence of confederates (primed subjects from the same group) who were simultaneously primed with the correct answer to the same task. We believe that the outcome of the primary group (dyads and adults) is statistically significant, rejecting the null hypothesis. The current research demonstrated inter-brain connectivity through ongoing emotions and motion dynamics, creating cooperation levels within individuals of the primary group beyond sensory modalities.

We did not expect the coherent intelligence effect in experiments with subjects of the secondary group; and we did not find it in the experiments with the SL3 (one experiment with 56 adults, M=21) and US3 tasks (2 experiments with 151 first-year students, M=19). Surprisingly, the outcome of the experiment No.12, 12/05/2020 with the task UL3 (an unfamiliar language) was very high; and the Chi-squared distribution and *P*-value of this experiment are very significant. This deviation from the expected result for this team was 380% higher than the probability of a random choice. We tested these subjects before the experiment on the knowledge of examined language, and the results of 4 students were excluded from the outcome, since they speak it. Other subjects did not speak this language at all. Following the secondary group definition, we rated this team as the secondary group, while they showed results as members of the primary group. The result of the experiment No.12 led us to additional study the history of these students' relationships and the formation of their team. Additional information at the stage of the results analysis showed that these second-year students had been visiting university facilities daily before the pandemic in 2020. They followed the same social rhythm during two academic years before the experiment. While studying at the university, they also completed an additional team training program to increase the level of inter-group cooperation. To our mind, the UL3 outcome means that these subjects had more close social cooperation between them than we thought at the moment of the experiment. This additional information may be useful for understanding why these subjects from the secondary group (as we thought earlier) also showed a significant accuracy in translating an unfamiliar language, as if they were members of the primary group. We believe that the results of the experiment No.12 (UL3) provides an opportunity for further research on the formation of close social bonds in groups that can also promote the more precise determination of the primary group's criteria. Further research on how the modality of communication can affect the message's understanding, even if the language of the message is unfamiliar to the recipient, can also develop knowledge about the inter-group cooperation.

The results of other teams with subjects of the secondary group in the tasks SL3 and US3 supported our expectations that there should be no coherent intelligence effect in the secondary group. The unexpected result in UL3 does not reject the significant outcomes of experiments in dyads (UL1, SL1, and US1) and adults from the primary group (UL2, SL2, and US2). We believe that the research outcome supports the hypothesis that from the beginning of cognition, the effect of coherent intelligence allows the nervous system to distinguish particular sensory stimuli from chaos, following shared intentionality with intimately related individuals. This outcome is consistent with our previous research outcome in 2019 from 7 experiments with 104 subjects, including 51 confederates and 53 participants [14]. The presented results complemented the already published research about 12 online experiments with 67 adults and children, *P*-value < 0,001 [19]. We believe that, future research should aim to understand the possible application of this effect to the advanced online curriculum for small children.

Authors Contribution Acknowledgment

Igor Val Danilov formulated the hypothesis and wrote the first draft of the manuscript. Igor Val Danilov and Sandra Mihailova improved the text over several iterations.

Acknowledgment

The authors express their gratitude to Prof. Iegor Reznikoff, Emeritus Professor Université de Paris Ouest Nanterre, France, for commenting this paper.

REFERENCES

1. J. Aschoff, "Comparative physiology: Diurnal rhythms," Annual. Rev. Physiol., vol. 25, pp. 581-600, 1963, doi: 10.1146/annurev.ph.25.030163.003053.
2. L. D. Grandin, L. B., Alloy, and L. Y. Abramson, "The social zeitgeber theory, Circadian rhythms, and mood disorders: Review and evaluation," Clinical Psychology Review, vol. 26, pp. 679–694. 2006, doi:10.1016/j.cpr.2006.07.001.
3. A. N. Meltzoff and M. K. Moore, "Imitation of facial and manual gestures by human neonates," Science, vol. 198, pp. 75–78. 1977, doi: 10.1126/science.198.4312.75.
4. I. Val. Danilov, "Imitation or early imitation: Towards the problem of primary data entry," Journal of Higher Education Theory and Practice, in press.
5. W. S. Condon and W. D. Ogston, "A Segmentation of behavior," Journal of Psychiatric Research, vol. 5, pp. 221–235, 1967.
6. G. Markova, T. Nguyen, S. Hoehl, "Neurobehavioral interpersonal synchrony in early development: The role of interactional rhythms," Front. Psychol. vol. 10, pp. 2078, 2019, doi: 10.3389/fpsyg.2019.02078.
7. D. O. Hebb, "Drives and the C.N.S." Psychol. Rev. vol. 62, pp. 243–254, 1955. doi: 10.1037/h0041823 .
8. G. E. Miller, E. Chen, and E. S. Zhou, "If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans," Psychol. Bull. vol. 133, pp. 25–45, 2007, doi: 10.1037/0033-2909.133.1.25 .
9. L. K. Cirelli, Z. B. Jurewicz, and S. E. Trehub, "Effects of maternal singing style on mother–infant arousal and behavior," Journal of Cognitive Neuroscience, pp. 1–8, 2019, doi: 10.1162/jocn_a_01402.
10. E. Prochazkova, and M. E. Kret, "Connecting minds and sharing emotions through mimicry: A neurocognitive model of emotional contagion," Neuroscience and Biobehavioral Reviews, vol. 80, pp. 99–114, 2017, http://dx.doi.org/10.1016/j.neubiorev.2017.05.013.
11. C. M. Heyes, "Empathy is not in our genes," Neuroscience and biobehavioral reviews, vol. 95, pp. 499–507, 2018. https://doi.org/10.1016/j.neubiorev.2018.11.001.
12. M. Katayama et al., "Emotional contagion from humans to dogs is facilitated by duration of ownership," Front. Psychol. vol.10, pp. 1678, 2019, doi: 10.3389/fpsyg.2019.01678 .
13. C. N. Cooley. Social organization: A study of the larger mind, New York, NY: Charles Scribner's Sons, 426 pp. 1909.
14. I. Val. Danilov, S. Mihailova, and V. Perepjolkina (2019). Unconscious social interaction, coherent intelligence in learning. Proceedings of the 12th annual conference ICERI Seville (Spain), pp. 2217-2222, doi: 10.21125/iceri.2019.0606.
15. I. Val. Danilov and S. Mihailova, "Emotions in e-learning: The review promotes advanced curriculum by studying social interaction," The proceedings of the 6th International Conference on Lifelong Education and Leadership for ALL, held at Sakarya University, July 2020. [retrieved: March, 2021]https://www.researchgate.net/publication/343053966_Emotions_in_e-Learning_The_Review_Promotes_Advanced_Curriculum_by_Studying_Social_Interaction.
16. M. L. Commons and S. J. Chen, "Advances in the model of hierarchical complexity (MHC)," Behavioral Development Bulletin, 19(4), 37-50, 2014.
17. I. Val. Danilov and S. Mihailov, "Knowledge sharing through social interaction: towards the problem of primary data entry." 11th Eurasian Conference on Language & Social Sciences which is held in Gjakova University, Kosovo. pp.226. [retrieved March, 2021] http://eclss.org/publicationsfordoi/abst11act8boo8k2021a.pdf. The Best Research Paper Award. 2021.
18. I. Val. Danilov and S. Mihailova, "Social interaction shapes infants' earliest links between language and cognition," Social Sciences Bulletin, 2:29, pp. 145-157, ISSN 1691-1881, eISSN 2592-8562, 2019.
19. I. Val. Danilov and S. Mihailova, "Ontogenesis of social interaction: Review of studies relevant to the fetal social behavior," In S. Denison., M. Mack, Y. Xu, & B.C. Armstrong (Eds.), Proceedings of the 42nd Annual Conference of the Cognitive Science Society (p. 3506). Cognitive Science Society. [retrieved March, 2021] https://cognitivesciencesociety.org/wp-content/uploads/2020/07/cogsci20_proceedings_final.pdf . 2020.

New Findings in Education:

Primary Data Entry in Shaping Intentionality and Cognition

Igor Val Danilov
Academic Center for Coherent Intelligence
ACCI
Rome, Italy
e-mail: igor_val.danilov@acci.center

Sandra Mihailova
Rīga Stradiņš University
RSU
Riga, Latvia
e-mail: sandra.mihailova@rsu.lv

Abstract— **One of the most intriguing questions in cognitive science is the Primary Data Entry (PDE) problem: where does social knowledge come from? After birth, an organism meets with reality, which is transcendental, staying beyond any experience and understanding of this pure reason. Basic approaches to the study of cognition are cognitivism, connectionism, and embodied dynamicism, The weakness of cognitivism and connectionism is the need to introduce an initial set of social phenomena of the specific community to trigger this system: the PDE problem. There is also a same gap in knowledge in embodied dynamicism approach, this interpretation of a dynamic system is not accurate. The current paper presents the Model of Coherent Intelligence (MCI) and its neural foundation. The analysis of recent empirical data proposes new insights on the origin of intentionality: (i) cognition begins from separating sensory stimuli: Long-Term Potentiation can be induced in neurons of particular Modality-Specific gateways (ignoring other stimuli)–selective induction promotes selective sensitivity to the chaos of stimuli. (ii) Neurons can learn spike-timing-dependent plasticity in social interaction: immature neurons learn the timing code to modulate certain synaptic strength, which triggers either Long-Term Potentiation or Long-Term Depression. The hypothesis of the MCI argues that social interaction shapes organisms' intentionality, promoting similar categorization of stimuli in intimately related individuals with shared social routine and interests. This approach provides a wide range of possibilities for developing a human-computer interface.**

Keywords-coherent intelligence; emotional contagion; interactional synchrony; social entrainment; embodied cognition

I. Introduction

The academic knowledge on the study of mind historically and conceptually has settled three main approaches within cognitive science: cognitivism, connectionism, and embodied dynamicism [1]. Many theories of mind combine all three approaches, where they co-exist in various hybrid forms. The more interesting of them are the Embodied dynamic system [1], the theory of innate intersubjectiivity and innate foundations of neonatal imitation [2], the theory of natural pedagogy [3], and the theory of sensitivities and expectations [4]. All these theories are plausible; the current paper observes different views to engage a gap in knowledge.

According to Thompson [1], cognitivism (the metaphor is the mind as digital computer) and connectionism (the mind as neural network), in different ways, appeal to the same computational principle of cognition. This principle based upon processing a signal within neural networks. This computational principle certainly requires the primary data entry as a necessary initial condition to launch processing. None algorithm and/or a sequence of instructions may perform the computation of any process without corresponding to the specific situation inputs, that should substitute variables and parameters of the formulas. The algorithm remains just a set of mathematical variables without this input. This argument may mean the necessity to input an initial set of social phenomena of the specific community to trigger this system – the Primary Data Entry (PDE) problem [5].

According to embodied cognitivists, the mind is an autonomous system by its self-organizing and self-controlling dynamics, which does not have inputs and outputs in the usual sense, and determines the cognitive domain in which it operates [1][6][7]. This approach is grounded on the dynamical hypothesis [8]. However, this interpretation of a dynamic system is not accurate. Why does the dynamic system need PDE:

Argument A. According this approach embodied features of cognition deeply depend upon characteristics of the physical body. If the agent's beyond-the-brain body plays a significant causal role, then the primary data yet makes sense.

Argument B. In mathematics, a dynamic systems model is a set of evolution equations. It means that entering primary data is required. The dynamic system may not begin its life cycle without introducing initial conditions corresponding to specific situation inputs and parameters.

Argument C. The dynamical system hypothesis [8] have not claimed the lack of initial conditions. Dynamicists track primary data less than dynamic changes inside. However, it does not mean that primary data do not exist and do not necessary.

Given these above arguments, the PDE problem must be considered in the onset of cognition. The embodied dynamic system approach tends to solve the above-noted gap by introducing the notion of dynamically embodied information [1]. Although, to introduce this concept, it is necessary to explain the categorization of reality through intentionality. According to embodied cognition approach, symbols encode the local topological properties of neuronal maps [1], a dynamic action pattern. The sensorimotor motor network yields pairing of the binary cue stimulus with the particular symbol saved in the structures and processes that embody meanings. 'Representational “vehicles” are temporally extended patterns of activity that can crisscross the brain-body-world boundaries, and the meanings or contents they embody are brought forth or enacted in the context of the system’s structural coupling with its environment [1, p.36].' This idea requires introducing the nature of intentionality. In a multi-stimuli environment, the stimulus-consequence pair is unpredictable due to the many stimuli claiming to be associated with the embodied dynamic information randomly. The bond of stimulus-consequence pair of a social phenomenon in the sensorimotor network requires categorizing reality by the nervous system before applying the innate reflex about this social phenomenon to a specific case. Therefore, dynamically embodied information can be useful if intentionality is already in place. Meanwhile, the embodied dynamic system introduces intentionality without a biological and / or physical basis. The theory of natural pedagogy [3], and the theory of sensitivities and expectations [4], as well as many others may not solve the problem of PDE [9].

According to Trevarthen and Delafield-Butt [10], primary consciousness develops in embryogenesis and is the first operative in early fetal life. 'Consciousness as "acting with knowledge" requires a nervous system that regulates prospective perception in intentional engagement with the world [10, p. 22]’. In the first trimester, patterns of sensory regulation of movements of the fetus' body and limbs gain affective evaluation and sensitivity for sounds and rhythms of other human presence [10]. It means that the pure nervous system should already possess intentionality as well as initial knowledge about social reality: human sounds and rhythms also yield meanings. Even if fetuses can hear different sounds and feel rhythms from outside of the womb, this does not mean that they alone (independently) can process their meanings.

Searle et al. [11] argued that intentionality is the mental power of minds to represent or symbolize things, properties, and states of affairs. According to Crane [12], mental states or events or processes which have objects in this sense are traditionally called 'intentional,' and 'intentionality' is, for this reason, the general term for this defining characteristic of thought. The meaning of directed action implies the purpose of the action, which first requires the categorization of reality. It's a dichotomy of what happens first. Current knowledge does not solve it.

Tomasello [13], through the study on ontogenesis and phylogenesis, introduced the hypothesis of gradually increasing social bond development in children referred to time slices: (1) emotion sharing from the birth, (2) joint intentionality from the nine-month revolution, (3) collective intentionality at around three years of age, (4) reason and responsibility. Tomasello [13] introduced the beginning of cognition through the newborns' basic motive force of sharing intentionality. However, the mechanism of such emotion coordination is not clear because it is grounded on emotion sharing [13]. Whether or not protoconversations imply understanding emotional states. Many researchers, including the authors, believe that the hypothesis about the universality of emotional expressions is formed by limited experimental methods, since other research designs show the opposite outcome [14]-[18]. There is no evidence of a genetic mechanism that can link meaning in mind with certain social reality to apply an appropriate emotional pattern to a specific situation. Even if one assumes that the hypothesis of universal emotional expressions proves innate emotional patterns together with their meanings; even if newborns may alone recognize the basic facial expressions of caregivers and the specific situation to apply them; but in this case, newborns do not have time for such a "training course", because they demonstrate their achievements already in the first hours of life [19]. If there is no innate mechanism, then, apparently, emotional contagion can occur between individuals without their awareness [9][19][21]; it can happen even without awareness of the emotional stimuli existence [21]. Section II discusses the hyperscanning and near-infrared spectroscopy studies' outcome, showing brain-to-brain synchronization. Section III presents, for the first time, the hypothesis of the neural foundation of shared intentionality. Section IV elaborates all findings, describing their meanings.

II. Problem: How Does Social Interaction Encourage Cognition

Brain-to-brain relationships shape the mind during moment-to-moment interactions [22]. The dichotomy of newborns' succeed in beginning knowing and their communicative disability challenges our knowledge on social interaction modalities [19]. We believe that understanding the problem of the intentionality emergence in an organism at the beginning can explain the problem of PDE and the onset of consciousness. This knowledge can contribute to the study of cognition because obviously if and as soon as this implicit modality occurs it continues the whole rest of life. We believe that the caregivers' intentionality forms the intentionality in newborns. Fetuses and newborns are not able to behave intentionally on their own due to the lack of meaningful (informative) sensory interaction at the beginning [5][19][23]. We predict an implicit modality of social interaction that provides shared intentionality at the beginning. Cooperation in a group enhances intentionality, providing categorization.

According to Valencia and Froese [22], their review of studies based on EEG- and fNIRS hyperscanning methodologies shows evidence of inter-brain synchronization in the fastest frequency bands, supporting the possibility of extended consciousness. Among hyperscanning studies, we have chosen 4 studies conducted without explicit interaction between subjects. These studies compared differences of brain-to-brain synchronization in subjects when participants solved tasks together as confronting to the condition in which: (i) the subjects solved them individually [24][25]; (ii) the same task when interacting with a machine [26]; (iii) the individuals from another team solved the same problem [27]. These studies declared an exclusion of sensory interaction between subjects. However, it should be noted that the subjects of all these studies knew about social encounters during the experiments. Therefore, instead of mental collaboration their results may simply mean an increase of brain activity due to similar emotional arousal in participants stimulated by the social encounter.

The near-infrared spectroscopy study (non-hyperscanning) on asleep newborns shows an increase of the neural response to a familiar (English language) versus unfamiliar language (Tagalog, a Filipino language) spoken by strangers in both conditions [28]. The language stimuli (the identical low-pass filtered sentences) were played through two speakers approximately 1.5 m from the infants' head. According to May et al., [28], this findings show that the newborn's neural processing of language is influenced by early language experience due to neonate brain responds to familiar versus unfamiliar language. To our mind, this outcome may lead to evidence of another inference. This experiment was not a hyperscanning technique. However, subjects were in pairs with their caregivers. Neonates classified these sound stimuli without the ability to perceive them. Sleeping newborns' brains reacted to sound stimuli that their sensing could not provide due to their brains' sensory isolation to meaningless and unfamiliar sounds. Sleepers seem to enter a standby mode, allowing them to balance the monitoring of their surroundings with sensory isolation [29]. Sleepers are sensitive to the semantic content of an auditory stream [29]-[31] and amplify relevant, meaningful stimuli [29][31]. The sleeping brain retains some residual information processing capacity, which, however, does not form enduring memories [32]. Neonates are not able to understand even Mother's speech although her sound is familiar. Given all these, any speech for neonates is meaningless, and asleep newborns may not be sensitive to the sounds even their native tongue (the language spoken by the mother during pregnancy) in experiments when these sounds were pronounced by outsiders. However, they were sensitive to them. Sleeping newborns' brains reacted to sound stimuli that their sensing could not provide due to their brains' sensory isolation to meaningless and unfamiliar sounds. We believe that this outcome may mean the implicit modality of newborns' interaction with caregivers since any other explanation of this outcome is excluded.

Whether or not social interaction contributes to the increase in results of a group mental activity? Whether or not dyads maintain non-perceptual interaction? What are the neurobiological grounds of intentionality?

III. Neural foundation of Coherent Intelligence

A. *Experiments on Problem-Solving in Groups*

Recent research of 24 online experiments presented that unprimed participants show a more significant accuracy level when they complete the thought task simultaneously with confederates who are primed with the correct answer; if they were emotionally stimulated and completed the tasks without communication [9]. Primary groups [33] show empirical evidence of a more significant accuracy in problem-solving in the coherent intelligence state. In specific, we conducted 13 experiments in dyads (116 subjects) with *P*-value < 0,001 (probability-value in null hypothesis significance testing), and 7 experiments in primary group adults (41 subjects) with the *P*-value < 0,002. Experiments with 43 secondary group subjects (unfamiliar adults, M=20) show the effect only with the task of unfamiliar language translation. Non-semantic tasks–with synthetic language and two-color round symbols–did not stimulate the effect in 2 experiments with 207 secondary group subjects (unfamiliar). These results are consistent with research Danilov et al. [34].

B. *The Model of Coherent Intelligence*

According to Danilov and Mihailova [23], a supranormal environmental case–e.g., first hours after birth–stimulates supranormal sensation in dyads. This can push the inherited mechanism of social entrainment of infants to the rhythm of the mother. Both the supranormal sensation and social entrainment may stimulate the common emotional arousal. The latter is increased by the ongoing supranormal sensation and the occurring rhythm of arbitrary movements of the infant. The continuing supranormal sensation and ever-increasing arousal of the infant and the mother along with the rhythm of the infant's unintentional movements stimulate early imitation and emotional contagion. The problem is how the infant capture and reproduce the kinematic of movements.

The MCI proposes that common emotional arousal together with the identical rhythm create coherent mental processes in dyads–Coherent Intelligence (Figure 1). At Sensorimotor Stage (by Piaget, or Stage 3 of the Model of Hierarchical Complexity MHC [35]), organisms do not maintain bilateral communication. According to Danilov and Mihailova [23], individuals are able to interact by distinguishing perceptual signals of identical modality by their significance. This ability can contribute to ostensive cues. After all, this meaningless interaction modifies into communication when individuals imbue perceptual impulses with mutually implied meanings, cascading their signals in response to the history of relations between them [23].

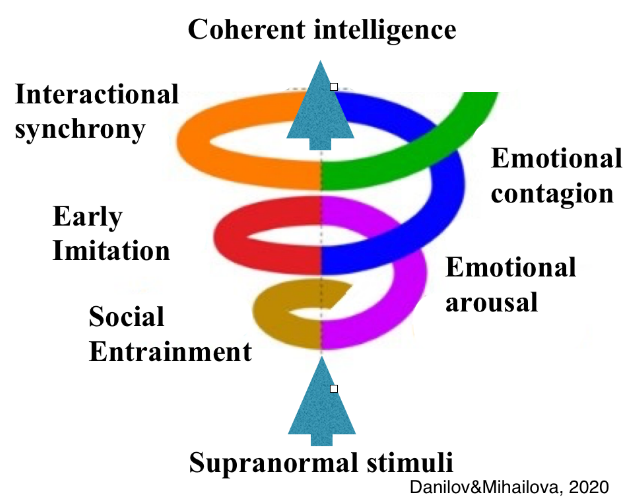


Figure 1. The Model of Coherent Intelligence.

C. Neural Foundation of this MCI

Leaving aside semantic constraints, it seems consistent to say that intentionality manifests itself in intention. Specific brain regions may be engaged in shared sensory/cognitive processes irrespective of the feedback's valence and in encoding the subjective relevance of the feedback [36][37].

Outside areas involved in this processing, additional brain areas are specifically engaged according to the particular communicative modality [38]. According to Tettamanti et al. [39], Intention Processing Network (IPN) involves the medial prefrontal cortex, precuneus, bilateral posterior superior temporal sulcus, and temporoparietal junctions. Depending on different social interaction modalities, the IPN is complemented by activation of additional brain areas, reflecting different Modality-Specific (M-S) input gateways to the IPN [39]. The M-S gateways mediate the structural and semantic decoding of stimuli and provide M-S information [39]. Sensory inputs of a specific modality can activate the precise association of certain sensorimotor networks with specific brain emotion circuits [40].

We believe that this emotion-motion coherence can involve the particular cognitive process of a high order. When two or more organisms are in common emotional arousal and simultaneously in the interactional synchrony, then these two different experiences may meet each other in high-order cognitive processing. Emotional arousal can trigger evolutionarily old brain circuits, which interact with high-order cognitive and linguistic processing [40]. It seems uncontroversial to say that infants' pure nervous system may experience emotions, but only primitive ones related to survival, such as those associated with hunger and pain. However, newborns cannot express emotions themselves appropriately to a specific social case on their own, even though they possess inherited neuronal patterns of primitive emotional impressions. They also cannot understand the expression of others' emotions (as is discussed above). They are only able of experiencing primitive emotions. Research on insects–organisms in stage 3 of MHC [35] like human newborns–assumes that they also experience emotions [41]. Researchers argued that agitated honeybees exhibit pessimistic cognitive biases: 'Whether animals experience human-like emotions is controversial and of immense societal concern. The next reason is that animals cannot provide subjective reports of how they feel, emotional state can only be inferred using physiological, cognitive, and behavioral measures. In humans, negative feelings are reliably correlated with pessimistic cognitive biases, defined as the increased expectation of bad outcomes. Recently, mammals and birds with poor welfare have also been found to display pessimistic-like decision making, but cognitive biases have not thus far been explored in invertebrates [41].'

In parallel, interactional synchrony stimulates a sensorimotor network engaging neural networks responsible for communicative intention processing (including high-order cognitive and linguistic processing)[38]. Neural networks of emotional excitation and the sensorimotor networks are separately connected to many different M-S gateways. Meanwhile their coherence intersects in certain M-S gateways of each organism depending on (i) pattern of neural circuit engaged through emotional excitation and (ii) pattern of the sensorimotor network [38].

We propose a rough hypothesis of how Long-Term Potentiation (LTP) can be induced only in particular M-S gateways, retaining information about the certain received stimulus. Different areas of the brain exhibit different forms of LTP, their types depend on a number of factors, such as age and the neuron's anatomic location. However, the common processes are the same for all. The simple nature of Hebbian learning, based only on the coincidence of pre- and post-synaptic activity, LTP is persistent, lasting from several minutes to many months, and it is this persistence that separates LTP from other forms of synaptic plasticity [42]. Spike-timing-dependent plasticity (STDP)–that involves the pairing of pre- and postsynaptic action potentials (APs)–causes a variation of LTP or Long-Term Depression (LTD) [43]. The timing between pre- and postsynaptic APs modulates synaptic strength, triggering LTP or LTD [43]. The sign and magnitude of the change in synaptic strength depend on the relative timing between spikes of two connected neurons (the pre- and postsynaptic neuron) [43]. The structural organization of excitatory inputs supporting STDP remains unknown [43]. Even though the ensemble of emotion-motion integrated networks weakly stimulates the intersected neurons in their junction with M-S gateways. If all M-S gateways also simultaneously receive weak stimulation from the receptors (due to the chaos of stimuli received by the pure nervous system), then this multi-signal contributes to LTP in the neurons of particular M-S gateway at the junction of this emotion-motion ensemble due to the effect of the synaptic cooperativity, because of the following. LTP can be induced either by strong tetanic stimulation of a single pathway to a synapse, or cooperatively via the weaker stimulation of many. Neurons from the gateways in the connections of these networks receive cooperative stimulation. Induction of cooperativity can ensure LTP.

According to Tazerart et al. [43], the synaptic cooperativity of only two neighboring synaptic inputs onto spines in the basal dendrites of L5 pyramidal neurons extends the pre–post timing window that can trigger potentiation. The engaged M-S gateways retain a certain stimulus, while other M-S gateways (also of the same sensory modality) remain depressed without keeping information of other stimuli. Therefore, specific M-S gateways are sensitive, and all these organisms respond to specific sensory modalities. Figure 2 shows a schematic picture of this process. The induced emotion and sensorimotor networks (they are red in the picture) together activate certain M-S gateway even with weak stimulation of sensory input. The different colors of M-S gateways refer to different sensory modalities. At this point, the analysis encounters the ground of the PDE problem of how immature neurons learn the timing code to modulate certain synaptic strength, which triggers either LTP or LTD. Because the structural organization of excitatory inputs supporting STDP remains unknown [43].

The study of the PDE problem leads to the analysis of the axiomatic foundations of Psychology, Sociology, and Neuroscience–the basic notions that form these sciences–from the perspectives of the actual scientific paradigm.

 ISBN: 978-1-61208-847-1

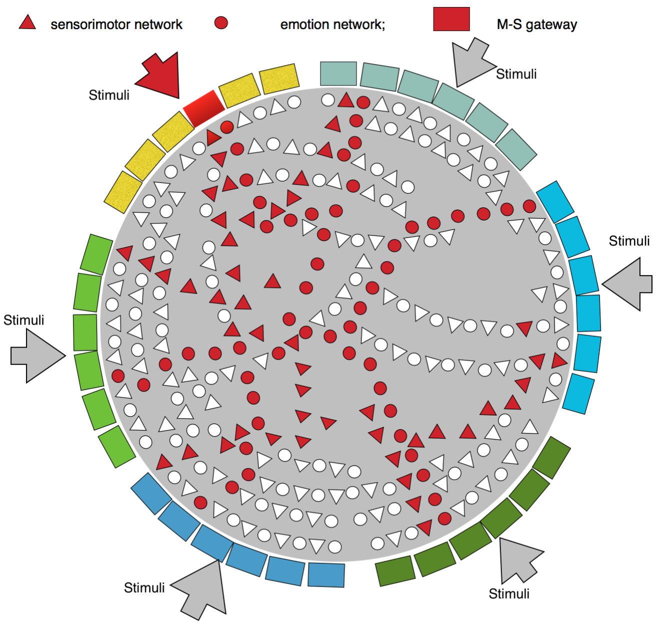


Figure 2. The Scheme of M-S Gateway Activation.

The question of "how can the blank mind begin to learn from social interaction" is reduced to "how immature neurons learn the timing code to modulate certain synaptic strength, that triggers either LTP or LTD". The sign and magnitude of the change in synaptic strength depend on the relative timing between spikes of two connected neurons (the pre- and postsynaptic neuron). How can neurons of an immature organism (even a newborn) learn the structural organization of excitatory inputs that support STDP? The further arguments show why we believe that the entanglement state of neurons can contribute to simultaneous LTP in neurons.

The daily routine develops neural patterns of primitive emotions and sensorimotor neural patterns in infants. Their everyday coherence with the social world forms various integrated neural patterns of different emotions from the existing ensemble of emotion scripts in their community. We believe that caregivers contribute to the formation of emotion scripts and, consequently, shaping of specific neural patterns in infants. Obviously, adults experienced intentionality before their coherent mental process began with newborns. Life experience taught them particular emotion scripts and defined their precise motion kinematics, that formed more elaborated sensorimotor patterns. In routine cooperation with newborns, a caregiver enters in interactional synchrony with a newborn, under the influence of supranormal stimuli, being in social entrainment. Therefore, the similar M-S gateways are excited in the dyad. Meanwhile, the adult's current intentionality has already triggered a particular network that includes current emotion patterns and sensorimotor patterns. Part of it corresponds to a primitive complex emotion-sensorimotor network in the newborn with similar M-S gateways. This newborn's primitive network is less developed, although it is similar to the part of the adult's integrated complex network. It can be assumed that the neurons in the connections of different excited emotion patterns and sensorimotor patterns go into the entangled state because they receive LTP, being induced cooperatively via many stimulations. In such a manner, the adult and infant neurons behave as a single unit. Therefore, specific M-S gateways are sensitive in dyads, and these organisms equally respond to specific sensory modalities. The induction of t-LTP and t-LTD in single spines follows a bidirectional Hebbian STDP learning rule [43]. Hebbian theory claims that an increase in synaptic efficacy arises from the learning process. The PDE problem in a chaos of stimuli requires a teaching mechanism from the beginning. The entanglement state of neurons is a possible option of how infants' neurons learn spike-timing-dependent plasticity. In the entanglement state, the behavior of the neurons of a mature organism determines and trains the neurons of a newborn.

Emotion sharing indicates implicit modality of social interaction. The entanglement state of neurons in the certain M-S gateways is a possible option of how infants' neurons learn STDP. This involvement of similar networks and the sensibility of the certain M-S gateways lasts as long as is necessary to teach the immature nervous system. The entanglement state of these neurons ensures their immediate response to the specific stimulus, regardless of the spatial division of organisms. Therefore, specific M-S gateways are sensitive, and these organisms equally respond to specific sensory modalities.

Knowledge about consciousness is being developed by studying the interaction of neurons, whose dimensions are similar to objects studied from the perspective of quantum mechanics scale. Knowledge about their behavior may complement the set of social interaction modalities. The generation of entanglement between increasingly macroscopic and disparate systems is an ongoing effort in quantum science [44]. Recent studies have shown that the behavior of objects 15 micrometers in size is consistent with the quantum world's laws, such as the phenomenon of quantum entanglement [45]-[47]. An entangled state was generated even between a millimeter-sized dielectric membrane and an ensemble of 109 atoms [44]. In comparison, a neuron's nucleus has a diameter of 3 to 18 micrometers, and a neuron has a size of 4 to 100 micrometers. The quantum entanglement state of particles is not something rare in Nature. Quantum systems can become entangled through various types of interactions, for the moment we know: (a) light's polarization. One of the most commonly used methods is spontaneous parametric down-conversion to generate a pair of photons entangled in polarization; (b) decay cascade of the bi-exciton. Fiber coupler to confine and mix photons, photons emitted from decay cascade of the bi-exciton in a quantum dot; (c) Hong-Ou-Mandel effect; (d) in the earliest tests of Bell's theorem, the entangled particles were generated using atomic cascades. An exciting question is how to create a quantum entanglement state of living cells. According to Marletto et al. [46], experiments show entanglement between living sulfur bacteria with quantized light. Whether or not neurons of different organisms become entangled through similar stimulation of their emotion networks?

Entanglement is an essential property of multipartite quantum systems, characterized by the inseparability of

 ISBN: 978-1-61208-847-1

quantum states of objects regardless of their spatial separation [44]. Recent study tested quantum entanglement over unprecedented distances, beaming entangled pairs of photons to three ground stations across China—each separated by more than 1200 kilometers. 'Yin et al. used the Micius satellite, which was launched last year and is equipped with a specialized quantum optical payload. They successfully demonstrated the satellite-based entanglement distribution to receiver stations separated by more than 1200 km [45]'.

We believe that these findings may contribute to an advanced curriculum, specifically in facilitating the teaching of young children.

IV. CONCLUSIONS

The analysis of recent empirical data yields a hypothesis of beginning cognition–the Model of Coherent Intelligence and its neural foundation of how the pure nervous system distinguishes sensory stimuli. This hypothesis postulates two new ideas of the PDE basis: (1) cognition begins from a separation of sensory stimuli: LTP can only be induced in neurons of particular M-S gateways (not all)–selective induction promotes selective sensitivity to the chaos of stimuli. (2) Neurons can learn STDP in social interaction by repeating the timing code of other organisms' mature neurons to modulate certain synaptic strength, which triggers either LTP or LTD. We believe that the MCI shapes intentionality in intimately related individuals. Coherent Intelligence is the integration of M-S gateways of particular brain areas, which contributes to different organisms' sensibility to similar sensory inputs. The 24 online experiments show empirical evidence of the MCI. We believe that this approach may contribute to study the mind and, in specific, to understanding the appearance of intentionality and the Primary Data Entry problem. We believe that these findings may contribute to an advanced curriculum, specifically in teaching 2- to 3-year-old children. Further research should also examine whether the MCI can provide a contactless interaction of the computer with neuronal circuits, in which the computer would become a part of extended mind. This approach provides a wide range of possibilities for developing a human-computer interface.

AUTHORS CONTRIBUTION

Igor Val Danilov formulated the hypothesis and wrote the first draft of the manuscript. Igor Val Danilov and Sandra Mihailova improved the text over several iterations.

ACKNOWLEDGMENT

The authors express their gratitude to Iegor Reznikoff, Emeritus Professor Université de Paris Ouest Nanterre, France, for commenting this paper.

REFERENCES

1. E. Thompson, Mind in Life : biology, phenomenology, and the sciences of mind. The Belknap press of Harvard University press. Cambridge, Massachusetts London, England. 1st Harvard University Press paperback edition. 2010.
2. J. Delafield-Butt and C. Trevarthen, Theories of the development of human communication. 2012. [Online]. Available from: https://strathprints.strath.ac.uk/39831/1/Delafield_Butt_Trevarthen_2012_Theories_of_the_development_of_human_communication_final_edit_060312.pdf [retrieved: March, 2021].
3. G. Csibra and G. Gergely, "Natural Pedagogy," Trends Cogn. Sci., vol. 13, pp. 148–153, 2009.
4. S. R. Waxman and E. M. Leddon, "Early Word-Learning and Conceptual Development". The Wiley-Blackwell Handbook of Childhood Cognitive Development. 2010. [Online] Available from: https://www.academia.edu/12821552/Early_Word-Learning_and_Conceptual_Development [retrieved: March, 2021].
5. I. Val. Danilov, "Imitation or Early Imitation: Towards the Problem of Primary Data Entry." The Journal of Higher Education Theory and Practice (JHETP). In press.
6. F. J. Varela, Principles of Biological Autonomy. ISBN-10:0135009502, ISBN-13:978-0135009505. 1979.
7. F. J. Varela and P. Bourgine, "Towards a practice of autonomous systems. In Towards a Practice of Autonomous Systems." The first European conference on Artificial Life, ed. F. J. Varela and P. Bourgine, pp. xi–xviii. Cambridge: MIT Press. 1992.
8. T. Van Gelder, "The dynamical hypothesis in cognitive science." Behavioral and Brain Sciences, vol. 21 (5), pp. 615-628. 1998. DOI: 10.1017/s0140525x98001733.
9. I. Val. Danilov and S. Mihailova, "Knowledge Sharing through Social Interaction: Towards the Problem of Primary Data Entry." The 11th Eurasian Conference on Language & Social Sciences which is held in Gjakova University, Kosovo. p.226. 2021. [Online] Available from http://eclss.org/publicationsfordoi/ abst11act8boo8k2021a.pdf. [retrieved: March, 2021].
10. J. Delafield-Butt and C. Trevarthen, Development of human consciousness. 2016. [Online]. Available from: https://strathprints.strath.ac.uk/id/eprint/57845. [retrieved: March, 2021].
11. J. R. Searle, W. S. Slusser, and M. Slusser, Intentionality: An Essay in the Philosophy of Mind. Cambridge University Press. 1983.
12. T. Crane, "Intentionalism." The Oxford Handbook to the Philosophy of Mind. Oxford: Oxford University Press. pp. 474–493. 2009.
13. M. Tomasello, Becoming human: A theory of ontogeny. Belknap Press of Harvard University Press. 2019. https://doi.org/10.4159/9780674988651.
14. R. E. Jack, "Culture and facial expressions of emotion". Visual Cognition vol. 21, pp. 1248–1286, 2013, https://doi.org/10.1080/13506285.2013.8353.
15. C. Crivelli and M. Gendron, In e Science of Facial Expression (eds José-Miguel Fernández-Dols & James A. Russell) Oxford University Press, 2017.
16. L. F. Barrett, R. Adolphs, S. Marsella, A. Martinez, and S. Pollak, "Emotional Expressions Reconsidered: Challenges to Inferring Emotion in Human Facial Movements." Psychological Science in the Public Interest, vol. 20, pp. 1–68. 2019. https:// doi.org/10.1177/1529100619832930.
17. K. Hoemann et al., "Context facilitates performance on a classic cross-cultural emotion perception task". Emotion. (Washington, DC). vol. 19, pp. 1292–1313. 2019.
18. M. Gendron, K. Hoemann, A. N. Crittenden, S. M. Mangola, G. A. Ruark, and L. F. Barret, "Emotion Perception in Hadza Hunter-Gatherers". Scientific reports, vol. 10, pp. 3867 , 2020. https://doi.org/10.1038/s41598-020-60257-2.

19. I. Val. Danilov, "Social Interaction in Knowledge Acquisition: Advanced Curriculum. Critical Review of Studies Relevant to Social Behavior of Infants". The Twelfth International Conference on Advanced Cognitive Technologies and Applications COGNITIVE2020. 2020.
20. J. Decety and P. L. Jackson, "The functional architecture of human empathy". Behavioral and Cognitive Neuroscience Reviews, vol. 3, pp. 71-100. 2004.
21. M. Tamietto et al., "Unseen facial and bodily expressions trigger fast emotional reactions". PNAS, vol. 106: pp. 17661-17666, 2009.
22. A. L. Valencia and T. Froese, "What binds us? Inter-brain neural synchronization and its implications for theories of human consciousness". Neuroscience of Consciousness, vol. 6(1), 2020. doi: 10.1093/nc/niaa010.
23. I. Val. Danilov and S. Mihailova, "Emotions in e-Learning: The Review Promotes Advanced Curriculum by Studying Social Interaction". The International Conference on Lifelong Education and Leadership for All (ICLEL). pp.8-17. ERIC Number: ED606507. 2020. [Online] Available from: https://faf348ef-5904-4b29-9cf9-98b675786628.filesusr.com/ugd/d546b1_2d77ecc9e07f4b0fb2cccce23af201f7.pdf. [retrieved: March, 2021].
24. C. Szymanski et al., "Teams on the same wavelength perform better: Inter-brain phase synchronization constitutes a neural substrate for social facilitation". Neuroimage, vol. 15, pp. 425–436, 2017.
25. F. A. Fishburn et al., "Putting our heads together: interpersonal neural synchronization as a biological mechanism for shared intentionality". Soc Cogn Affect Neurosci, vol. 13, pp. 841–849, 2018.
26. Y. Hu et al., "Inter-brain synchrony and cooperation context in interactive decision making". Biol Psychol, vol. 133, pp. 54–62, 2018.
27. L. Astolfi et al., "Neuroelectrical hyperscanning measures simultaneous brain activity in humans". Brain Topogr, vol. 23, pp. 243–256, 2010.
28. L. May, K. Byers-Heinlein, J. Gervain, and J. F. Werker, "Language and the newborn brain: does prenatal language experience shape the neonate neural response to speech?" Front. Psychology, vol. 2, pp. 222, 2011, doi: 10.3389/fpsyg.2011.00222.
29. T. Andrillon and S. Kouider, "The vigilant sleeper: neural mechanisms of sensory (de)coupling during sleep." Current Opinion in Physiology, vol. 15, pp. 47–59, 2020.
30. A. Ibanez, V. Lopez, and C. Cornejo, "ERPs and contextual semantic discrimination: degrees of congruence in wakefulness and sleep." Brain Lang, vol. 98, pp. 264-275, 2006.
31. G. Legendre, T. Andrillon, M. Koroma, and S. Kouider, "Sleepers track informative speech in a multitalker environment." Nat Hum Behav, vol. 3, pp. 274-283, 2019.
32. R. Cox, I. Korjoukov, M. de Boer, and L. M. Talamini, "Sound Asleep: Processing and Retention of Slow Oscillation Phase-Targeted Stimuli." PLoS ONE, vol. 9(7), pp. e101567. 2014, https://doi.org/10.1371/journal.pone.0101567.
33. C. H. Cooley, Social organization: A study of the larger mind , (pp. 23-31). New York, NY: Charles Scribner's Sons, xvii, pp. 426, 1909.
34. I. Val. Danilov, S. Mihailova, and V. Perepjolkina, "Unconscious social interaction, coherent intelligence in learning." The 12th annual conference. ICERI 2019. doi: 10.21125/iceri.2019.0606.
35. M. L. Commons, "The fundamental issues with behavioral development." Behavioral Development Bulletin, vol. 21(1), pp. 1-12, 2016, http://dx.doi.org/10.1037/bdb0000022.
36. S. Oldham, C. Murawski, A. Fornito, G. Youssef, M. Yucel, and V. Lorenzetti, "The anticipation and outcome phases of reward and loss processing: A neuroimaging meta-analysis of the monetary incentive delay task." Human Brain Mapping, vol. 39, pp. 3398-3418, 2018.
37. D. Martinsa et al., "Mapping social reward and punishment processing in the human brain: A voxel-based meta-analysis of neuroimaging findings using the Social Incentive Delay task." 2021, bioRxiv preprint doi: https://doi.org/10.1101/2020.05.28.121475.
38. I. Enrici, M. Adenzato, S. Cappa, B. G. Bara, and M. Tettamanti, "Intention Processing in Communication: A Common Brain Network for Language and Gestures." Journal of Cognitive Neuroscience, volume 23, issue 9, pp. 2415-2431, 2011, https://doi.org/10.1162/jocn.2010.21594.
39. M. Tettamanti, M. M. Vaghi, B. G. Bara, S. F. Cappa, I. Enrici, and M. Adenzato, "Effective connectivity gateways to the Theory of Mind network in processing communicative intention." NeuroImage, vol. 155, pp. 169-176, 2017, doi: 10.1016/ j.neuroimage.2017.04.050.
40. L. Nummenmaa et al., "Emotional speech synchronizes brains across listeners and engages large-scale dynamic brain networks." NeuroImage, vol. 102, pp. 498-509, 2014.
41. M. Bateson et al., "Agitated honeybees exhibit pessimistic cognitive biases." Curr Biol., vol. 21;21(12), pp. 1070-1073, 2011, doi:10.1016/ j.cub.2011.05.017.
42. W. C. Abraham, "How long will long-term potentiation last?". Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. vol. 358 (1432), pp. 735–744, 2003, doi:10.1098/rstb.2002.1222.
43. S. Tazerart et al., "A spike-timing-dependent plasticity rule for dendritic spines." Nat Commun vol. 11, pp. 4276, 2020, https://doi.org/10.1038/ s41467-020-17861-7.
44. R. A. Thomas et al., "Entanglement between distant macroscopic mechanical and spin systems." Nat. Phys., 2020, https://doi.org/10.1038/ s41567-020-1031-5.
45. J. Yin et al., "Satellite-based entanglement distribution over 1200 kilometers" [Online] Science: vol. 356, issue 6343, pp. 1140-1144, doi:10.1126/science.aan3211, 2017. [retrieved: March, 2021].
46. C. Marletto, D. M. Coles, T. Farrow, and V. Vedral, "Entanglement between living bacteria and quantized light witnessed by Rabi splitting." Journal of Physics Communications, vol. 2(10). [Online] Available from: http://iopscience.iop.org/article/10.1088/2399-6528/aae224/ meta. [retrieved: March, 2021].
47. M. Sillanpää and S. Hong, "Spooky quantum entanglement goes big in new experiments." [Online] Science News 25 April 2018. Available from: https:// www.sciencenews.org/article/spooky-quantum-entanglement-goes-big-new-experiments. [retrieved: March, 2021].

 ISBN: 978-1-61208-847-1

The Visual Consciousness Space:

A mathematical proof of the irreducibility of consciousness to physical data.

Iegor Reznikoff
Département de Philosophie
Université de Paris Ouest Nanterre
Paris, France
e-mail: dominiqueleconte@yahoo.fr

***Abstract*—We show here that what we call 'visual space of consciousness', the space of what we see, is a specific space different from the purely physical one and that its properties imply that it cannot be reduced to or deduced from physical laws. Some biological points are also briefly considered. The arguments are of logical, mathematical and physical character; although elementary they require a careful reading (A first shorter version of this paper appeared in a hardly accessible Journal [1], and presented at the International Congress of Logic, Methodology and Philosophy of Sciences, in Beijing, August 2007). There is no need to define consciousness; we only observe some of its properties, namely geometric and topological properties of visual consciousness, and show that these properties cannot be based on physics only. Now, if a part of consciousness cannot be grounded on physics only, it is the same for consciousness as a whole and we speak of the irreducibility of consciousness to physical data. We do not consider philosophical questions or issues; in a simple physical and mathematical frame we give a logical proof of this irreducibility. Elements for a formal mathematical, logical proof are mentioned at the end of the paper. The idea of non reducibility of consciousness to physical properties creates a theoretical foundation of a new approach to study possibilities of contact-less brain-computer interface.**

Keywords-consciousness; vision; physics; reality; continuity; logic.

I. INTRODUCTION

The main purpose of this work is to give a proof of the non reducibility of consciousness to physical data. In order to treat the problem precisely and have clear definitions we essentially limit the question to the *visual space* i.e., to the space we see (when looking at something), so that we need not define *consciousness*. If a part – the visual one – of consciousness cannot be founded on physics only, it is the same for consciousness as a whole and we speak of the irreducibility of consciousness to physical data. In Sections II and III, notions of *visual space* and *irreducibility* are respectively defined; then we study two main properties of the visual space, namely its *continuity* (Section IV) and *unity* (Section VI). Since regarding unity the biological level is concerned, this point is briefly discussed (Section VII). The last Section gives elements for a formal mathematical and logical proof (which is out of the scope of this paper). Finally, we conclude with a short historical perspective. The strictly scientific approach of this work gives possibly new ways to study the contact-less brain and computer interface.

II. THE CONSCIOUSNESS SPACE

For a given observer, let A be the space of 'physical reality' as known by physics, the 'real' space of matter with what is included in it: moving atoms, particles and waves; and let B be the observer's brain regarded as a space, with its physiological and neuronal activity (of course $B \subset A$); then let C be the space of the observer's perceptive consciousness: what he sees, hears, touches, etc. considered as a space. There are of course further levels of consciousness, in particular a witness consciousness: the one that sees, hears, etc (not to speak of the thinking one). But here we consider only perceptive consciousness, what is seen (heard, etc), and, more precisely, the visual space in the case of vision (resp. the spaces of what one hears, touches, etc. for the other senses). For simplicity and because of its obvious geometric appearance, we confine our remarks mostly to the visual space, but the same points can be made about other spaces of perception (For a very interesting approach of consciousness of sound, see [2] and [3]). In what follows we will speak of A, B, and C also as being respectively the '*real'* or *physical space*, the brain space and the *consciousness space* with its visual sub-space.

Between a part of A – B (elements of A that are not in B) and B there is a map, say f, which to physical events in this part of A – B, through the perceptive channels, associates reactions in the brain space B. For instance, a photon flux received by the eyes creates an activity in the optical nerve and then in the brain. To this activity at the level B corresponds in general a representation in C; let g be this correspondence between brain activity and its representation (image) in the consciousness space. There is, therefore, a correspondence from A to C defined by $h= g\,[f]$. We have the following diagram (Figure 1).

If we limit C to perceptive consciousness – as we are doing here – we could expect to have $g\,[f(A - B)] = C$, but this is not the case since we are going to show that there are properties of C which do not proceed from properties of A. Note that f is not injective, since different points in A cannot always be distinguished in B, and even less in C. One might discuss, of course, the precise domains of definition of f and g – since there are not, for all events in A or in B, corresponding reactions or representations in B or C, respectively, but this is of no importance here. Also we

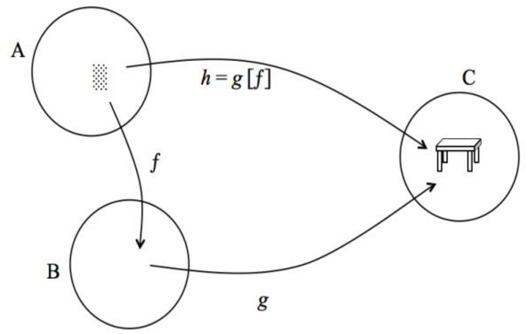


Figure 1. Spaces A, B, C as presented in the text.

did not give a precise mathematical definition of the maps since it is not even clear what the elements in A, B and of course in C are; one should probably rather speak of a map between some subsets of A and subsets of C, or even better in the language of categories; for an approach of some properties of consciousness in terms of categories, see [4] (I am thankful to the mathematician G. Choquet who mentioned this remarkable work). For sake of simplicity, we keep the elementary formulation in the language of set theory, also because we shall have a logical approach referring to some axioms of set theory. The given correspondences f and g are however clear enough to say that h gives a representation or an image of the reality A in the consciousness space C. For instance: to a subset T of A – B, a subset of particles reflecting a flux of photons, is associated, through the brain space B, a representation h(T) of T, in C, say a table; this table is a representation of external physical data (the set of all particles, waves, etc. concentrated in the given space we call *table*). We are going to study some properties of the space of such representations, what we call the consciousness space C.

Proposition 1. The space C is a specific space in itself, different from A and B.

Proof. We have C ≠ A, since the representation of A – B in C depends on B (and g), whereas A – B does not. For example if B is disturbed or injured, clearly so will be C (for instance a person sees two tables or none), while A – B has not changed (there is still one table). But also C ≠ B. What we see is not the brain, nor the activity of the brain. The same arguments show, moreover, that $C \cap A = \varnothing$ and $C \cap B = \varnothing$. Stated otherwise, the proposition says that

the consciousness space is different from the real physical space and from the brain space.

Given an object X in A, we do not see X in itself but only the result in C of a physiological and neuronal activity in B created by photons coming from X. This is well known, but the unconscious identification of h(X) with X, of what we see with the material physical world, is so strong and so widespread – even for those who have read Kant – that it is necessary to restate clearly: there is a proper space of consciousness and the picture that one sees is not the physical world. It is merely a representation of a set of particles and waves in the consciousness space; it does not mean, however, that this space is itself a 'set of particles': this is the point, the visual space is a proper non discrete space (see below).

An interesting question is that of the dimensionality of the consciousness space. For the visual space, one usually counts three spatial dimensions and three dimensions for colours. But we do not see a mixture of three colours; we see at least seven colours and their varieties and mixtures (For a topological approach of visual and colour spaces, see [5], [6], [7]). And what about the dimensions of the spaces of what we hear [3], smell or touch? Here, however we do not consider this question.

III. On irreducible Properties of the Consciousness Space

It is clear that the consciousness space has properties that are not as such in the physical space: this is the case, for instance, for colours, which are indefinable without the direct experience of seeing them (a person blind from birth can have no idea of what *green* means, though he might associate other feelings with this word). However, although the qualities (*qualias*) of a colour cannot be given physically, they do correspond to physical data: a green object reflects the light waves with a frequency that we see as 'green'. We say that this property of being green can be *reduced* to physical properties.

The question then arises whether there are properties of the consciousness space that cannot be reduced to real physical ones. If so, we speak of *irreducible* properties.

IV. Continuity of the Consciousness Space

A remarkable property of the visual space (but it is true also of other perceptions, although the matter is more difficult to formulate) is its continuity. We understand continuity in the mathematical sense, but the following elementary definition is sufficient here: in the visual space, there are no gaps or moving separate points; e.g., an ordinary white sheet of paper appears uniformly, permanently white and still, for at least a while. In contrast, physical reality at the atomic level is essentially discrete, non uniform, never motionless, and full of collisions; it doesn't mean that there are holes of energy or whatever, but we number atoms, electrons and various particles. With modern laser and other technologies, an isolated electron can be observed [8], [9], and with the NV (Nitrogen Vacancy) nanotechnology it is possible to produce sources of isolated photons [10]. Moreover, particles and waves are in perpetual movement. Of course, at our macroscopic level, we can use a magnifying glass and discover other aspects of what our eyes and consciousness did not see before, but the image remains continuous; there are no holes in the space of visual consciousness. Thus we have the question: how can a discrete moving atomic reality be represented in a continuous way (both in space and time)? The usual explanations about such questions concern the macroscopic level of perception. A typical example is given by the continuous appearance of a discrete pile of stones seen from a distance: it seems to be a white continuous spot. The usual argument is that the discreteness is too subtle to be perceived. The question then arises: to be perceived by what? And where? Indeed this gives no explanation at the atomic level, since the perceptions are transmitted and received (if we remain in a purely physical world) by discrete processes of particles and moving waves, particularly photons, and charges. How does this produce a continuous image, and where does this image appear? Certainly not in a physical space of particles, nor in the neuronal brain; the neurones transmit physical information up to consciousness which produces the continuous image we see; this is also a proof that C in no way belongs to the physical or brain space, and is a specific space of non-material, non-physical nature. Something to be perceived needs a perceiver and here the perceiver's visual continuous space cannot be reduced to physics because of the argument above. If everything were created, transmitted and received by physical spaces, it would remain permanently moving and discrete. As we say in French, 'the most beautiful girl in the world cannot give more than what she has': physics gives no more than physics.

Proposition 2. The property of continuity of the consciousness space C is not reducible to or even explainable in terms of physical reality.

Against this statement, there is also the argument that if there were holes in the visual space we simply would not see them, since we obviously can't see what we do not see (!). The discrete structure therefore cannot be perceived. However this argument again supposes that something is perceived and already presupposes a perceiver: who is the *we* in the sentence above and who or what is seeing? In such answers it is assumed that something (somebody) already sees or doesn't see; and the question remains how, in the final analysis, a 'continuous' space of vision can exist and where it can be located. Since it cannot be based on physics only, the conclusion is straightforward: continuity is a creation of consciousness. And here we come to a purely mathematical and logical consideration: continuity is not definable from discreteness and finite considerations, and cannot exist in a finite numbered domain. But physical reality – in a bounded domain at least – is finite. The property of continuity is, therefore, indeed irreducible to any physical reality, unless the notion of continuous field be introduced, which is a very theoretical and problematic notion that we discuss below.

A. Commentary

One can discuss whether the property of continuity is needed to characterize the visual space; for instance, isn't the property of *density* sufficient (as for the line of rational numbers)? Let us recall that density means that between two points there is always a third one. The above irreducibility argument remains in force even if we assume density; since density implies infinity, even in an interval or bounded space; indeed the fact that between two points there is a third one implies that between these two there is an infinity of points. An absolute proof of the continuity of the visual space in the strong mathematical meaning is certainly not possible because it requires high technical considerations of infinite character – let us recall that the continuity of a space implies that its infinity is not countable, which means that it is bigger than the infinity of the set of integers (it is said to have the power of the continuum). Such considerations are purely theoretical and certainly beyond any experience. But there is another strong epistemological argument for attributing continuity to the visual space. This argument comes from answering the question: how did the concepts of geometrical (Euclidian) space and precisely of continuity appear? How the geometrical line was and is understood to be continuous?

The notions of *geometrical space* and line appear of course in and from our visual space and visual experience (connected with that of movement and touch for three-dimensional awareness). Moreover, all our intuition of space geometry in the plane comes essentially from our visual space which, as we know, until the discoveries of Relativity Theory, was considered in its Euclidian formulation to be absolute (from the physical point of view at least). The notion of *continuity* proceeds as well from our visual experience, the best notion of a continuous line or surface being probably given by a surface of water: there are no holes or separations. That a segment of a straight line has infinitely many points (because it is dense) is readily understood and has been understood since Antiquity as well as the (intuitive) continuity of the line. And it is most remarkable that children, from visual experience, easily understand the notion of a (straight) line as well as its potential infinity and its continuity as being with 'points everywhere' so that there are no holes left. Of course the notions of closeness, or of going through are also related to our experience of movement and touch, but, finally it is by reasoning on the geometric line, which belongs to and comes from our visual space (so that it can be drawn), that the theory of this geometric line has been worked out. Also let us note that in our visual space all (necessary macroscopic) movements are continuous: it is impossible to join two points without passing somewhere in-between, while this is not the case at the atomic quanta level. It is therefore quite reasonable to consider that our intuition and understanding of the visual space *demonstrate it to be a continuous space*.

What we said about spatial continuity can be repeated concerning the continuity of the visual space in time (and, more generally, of perception in time). The visual space lasts in a continuous way as does a continuous movement; while at the atomic level, in duration, there is no continuity at all. However our perception of time is continuous and has led to a theoretical treatment of time which identifies it with the geometric line (for a study on time based on a distinction of physical and mind levels, see [11]). This continuity in time is closely related to what appears to be an even more remarkable property of consciousness and particularly of the visual space, namely its unity (see below, Section VI).

B. The consciousness space as a field

The only physical approach to continuity is given by the notion of field. For instance, an electromagnetic or gravitational field is assumed to be defined and active everywhere in the physical domain where it acts, and this *everywhere* is understood to be continuous since the space where the field is active is mathematically considered to be the three-dimensional space R3. This is a purely mathematical and theoretical formulation: we can only verify that the field acts on every particle or object appearing in the domain, and experience can show no more. But to assume continuity and R3 allows us to use the mathematical infinitesimal calculus with all its tremendous power. However, we claim, after the discussion above, that this geometrical approach is a creation of consciousness and particularly of consciousness of the visual space, since there is no other evidence for such a geometrical and topological continuous conception. It is not and cannot be given by direct physical experience which is finite. For us, therefore, this geometry is not in A, but in C, and then induced from C to the theoretical, mathematical treatment by using the space R3 containing a theoretical model of A.

But since this notion of continuous field actually exists in physics – be it created by and conceived from visual consciousness – we may say that consciousness is indeed a field. And just as a movement of electrons creates an electromagnetic field, we may conjecture that intense brain activity – of purely physical character at the atomic level of particles and waves – creates a field of consciousness: the greater the brain activity, the richer the field of consciousness. This field is, of course, not physical, since $A \cap C = \emptyset$ as we have seen, the space C being a specific one. Nor it is simply reducible to a known physical field, since in visual consciousness, we can isolate forms, colours, objects, etc., while even if there are different wave lengths, etc. in physical fields, it is consciousness that extracts the mentioned properties or elements from the visual space we see. There is nothing analogous or even expressible concerning physical fields. The property of seeing separate objects in the unity of the whole visual picture corresponds to the *Comprehension Axiom* of set theory: given a set E and a property P, the subset of elements of E verifying the property P, exists. It cannot be deduced from physics. It seems to be a fundamental property of consciousness related to the *a priori* capacity of consciousness to pay attention. Moreover, if we consider the whole perceptive consciousness, there is no homogeneity between visual pictures, acoustic perception or touch.

If we consider the spaces of what we see, hear, smell, taste or touch, as different subspaces of the whole consciousness space, the non-homogeneity of these subspaces is a quite peculiar fact, bearing in mind that they are all produced by the same kind of neuronal activity, since there seems to be no difference between the neurones of different perception areas in the brain. How can the same kind of neuronal activity produce such different worlds of perception? This question could yield another proof of the irreducibility of the consciousness space to the physical one. Of course the scales of various physical data producing the perceptions are quite different e.g., the scales of light waves, molecules (for the smell) or sound waves, but this does not explain the complete non-homogeneity of the corresponding subspaces of consciousness, whereas they are held together in a remarkable unity: I smell the rose that I see. This is a specific property of consciousness.

But since the emergence of space consciousness comes mostly from an intense brain activity of quantum electromagnetic nature, the relationship between such a field and the field of consciousness has to be investigated not only for isolated phenomena (for instance the fact that different frequencies of light produce different colours in the visual consciousness) but in the whole. Why would not a special intense physical activity – in the brain – create a field of different – non-physical – nature? Clearly, the intermediate biological level appears as an essential one.

V. The Objectivity of the Consciousness Space C

A peculiarity of the consciousness space is that it can be studied essentially only from inside, by itself: only consciousness knows consciousness. And "if you want to know my consciousness, look in yours" sounds as a wise saying. Thus we come to this important statement:

Proposition 3. The properties of C can be seen by everybody: its study is therefore perfectly objective.

Note that here the word objective has the same meaning as in natural sciences, e.g., physics, since everything we know is known from our perceptive consciousness, and everything we look at – for instance the position of a needle in a measuring apparatus – is seen in our visual consciousness, that is in C. If two persons see the same object (the needle at a given position) it is because it is the 'same' object in their respective visual spaces. Although the meaning of the word *same* cannot be explained; this meaning is based on a universal understanding without which no communication would be possible. Here, appears the common but meaningless question whether we all really see the same colours or objects: is the red that I see the same as the one you see? The question is meaningless because it cannot be verified, but *the simpler the hypothesis, the better it is*, and the simplest is to consider that indeed we have essentially the same consciousness. But as mentioned above, philosophical discussions are not considered here.

VI. Unity of the Visual Consciousness Space

One of the most remarkable properties of consciousness space – and moreover difficult to understand – is its unity, that is the capacity that consciousness has to gather perceptions as a whole; from a multiplicity of independent nervous impulses and neuronal processes consciousness produces a unified whole. We do not have consciousness of separated elements, but always of a coherent whole, even when looking at an isolated object.

This unity principle is the following: given separate elements $x_1, \ldots, x_n$, it is the actual capacity to conceive their totality i.e. the set $\{x_1, \ldots, x_n\}$. It is remarkable that this corresponds to an axiom of set theory; logically this property is not reducible. It cannot be deduced simply from the existence of $x_1, \ldots, x_n$ as separate elements. Therefore, it is not physically explicable, unless of course it is implicitly assumed (which is often the case, for instance when one assumes that things are somehow and somewhere 'observed' before any consciousness has been introduced). In particular, the argument that unity results simply from the simultaneity of neuronal processes in some centre of the brain is doubly inconsistent. First because the notion of simultaneity is meaningless without the notion of *now* or the notion of *at the same time as*, which presupposes a reference and a clock and therefore an observer, i.e. a consciousness that grasps this simultaneity, this very notion introduces already an observer, it is not an absolute notion. And the second inconsistency is that simultaneity presupposes certainly the comparison of at least two elements and hence the notion of totality, be it only of the set $\{x_1, x_2\}$ as a whole. Therefore, to have the notion of simultaneity we already need that of unity; it is impossible to avoid circularity. The simultaneity of physical events is perhaps necessary for consciousness of unity but not sufficient to explain it.

But even if the notion of simultaneity is given, the probability that all the possible visible 'dots' of our visual neurology (e.g., retina) are grasped together in a coherent unity (their number can be estimated of the order of 10^7), this probability is of order of $2^{(10^7)}$ (2 to the power of 10 to the power 7), which is well beyond any physical meaning even at the level of light-wave length. Unity cannot emerge 'by chance'; moreover, it is permanent, continuous in time. The probability for this continuing unity is physically without meaning.

This capacity of totalization, this gift of perceptive consciousness, is certainly one of its most important properties and unity may be the most characteristic property of consciousness. Consciousness unifies elements that otherwise are not related; from this comes what is called *meaning*. But we are not discussing this here any further.

As we have seen, the unity principle has, of course, no equivalent in physics; theoretically, it has to be borrowed from logic. The property of unity, say of visual consciousness and of the space C, is thus irreducible.

Proposition 4. The unity of consciousness is not reducible to physical properties.

The question then arises how far logical arguments can be used in physics, biology and matters of consciousness. But if one looks at a deductive science, rigorously founded, the logical and mathematical arguments are hitherto unavoidable. Of course a science can be very rich as a descriptive one, but the claim is now: is it possible to 'explain consciousness' by neuronal and finally from purely physical processes? Since in our attempt, *explain* means *deduce* or reduce, the argument needs to avoid circularity, therefore a careful logical examination is needed, which we have attempted above: the unity of the visual space cannot be reduced to or explained by physics without circularity since a notion of unity is needed beforehand.

VII. Biological Unity

If at some level the property of unity is needed and has to be introduced as such, then, it could be given already at a different level. It is natural to assume this unity, as we have seen, as one of the characteristic properties of consciousness, but it could be attributed already at the biological level. One often speaks of the 'unity of the cell'. Is it not at this elementary biological level that a principle of irreducible unity has to appear?

That such unity is necessary as a global *principle* in biology is simply shown by the same argument as the one given for the impossibility of a random unity of the visual space. Suppose a biological organism of about 10^9 (of about 10 to the power of 9) components (e.g., molecules); the probability that all these components should behave *together* in the right way in order to constitute a biologically viable unity, this probability is at *least* of the order of $2^{(10^9)}$ (2 to the power of 10 to the power of 9), which is, as we have seen before, beyond any physical meaning even at the atomic level: the age of the universe would not be sufficient for even one cell to have a chance to exist, not to speak of a more complex organism.

However, even if a biological property should normally appear and be stated before properties of consciousness (and moreover could explain some of its aspects), we have a knowledge of the visual space, of its continuity and unity, certainly clearer, at least in its immediacy, than an as yet unformulated principle of unity in biology.

VIII. Formal Approach

For a formal, strictly deductive logical approach, we need different levels of axioms, laws and data, so that the following levels have to be distinguished.

1. The Logical level needed for mathematics. This introduces axioms e.g., the *Axiom of Totality*: given $x_1, \ldots, x_n$ the set $\{x_1, \ldots, x_n\}$ exists. The *Comprehension Axiom*: given a set E and a property P, the subset of elements of E verifying the property P, exists. And finally an *Axiom of Infinity*.
2. The Mathematical level: theory of Real Numbers and Analysis. Logical axioms are intended for mathematical notions and reasoning.
3. The Physical level with its proper axioms and laws. The notions of continuity etc. are borrowed from the level 2.
4. The Biological level (this level is not really needed here).
5. The Consciousness level.

It is important to stress that we need not define consciousness (which would be a big challenge since consciousness is irreducible to other levels); we only observe some properties of visual consciousness, i.e. of what we see. But for continuity, unity and consciousness of seeing various objects, we need axioms analogous to the axioms above; these axioms are necessary to explain the mentioned properties of the visual space, and necessary for a deductive construction showing rigorously the irreducibility of consciousness to other levels. Since these axioms are not given by physics, clearly the level of consciousness is not at the physical level and cannot be deduced from physics. The careful scientific approach presented here may open new possibilities for studies concerning a contact-less brain and computer interface [18].

IX. CONCLUSION

That consciousness space is relatively independent from external physical reality is a classical statement. For Plato, Consciousness precedes Matter (as we learn from the *Timaeus*, 34c), the same for Indian classical religious philosophy; Kant's thoughts on this topic are well known (*Kritik der reinen Vernunft*), but it is worth quoting Berkeley: "*The proper objects of sight not without the mind; nor the images of anything without the mind*" and also "*Images in the eye are not pictures of external objects*" [12]. Here, we have simply shown that this relative independence of the consciousness space and its specific nature can be proved convincingly.

REFERENCES

[1] I. Reznikoff, "The consciousness space," Bio-Math, tome XXXX (n° 158), Paris, pp. 1-7, 2003.

[2] R. Casati and J. Dokic, La Philosophie du Son, Nîmes (France), 1994.

[3] I. Reznikoff, "On primitive elements of musical meaning," http://www.musicandmeaning.net/issues/showArticle.php?artID=3.2, 2005.

[4] A. Ehresmann, http://perso.wanadoo.fr/ vbm-ehr/.

[5] C. Pélissier, Aspects physico-mathématiques de la vision de la couleur, Thèse de Doctorat d'Etat, Université Paris VI, 1982.

[6] C. Pélissier, "Théorie unifiée de Young – Helmoltz – Hering, topologie induite, phénomènes d'adaptation, notion de temps et d'énergie" in Actes du 5e Congrès de l'Association Internationale de la Couleur, Mondial Couleur, Monte-Carlo, 1985.

[7] C. Pélissier, Couleurs et Temps, de la Physique à la Phénoménologie, Paris, 2006.

[8] D. Wineland et al., Phys. Rev. Lett., 31, 1279, 1973.

[9] A. Aspect et al., Phys. Rev. Lett, 49, 91, 1982.

[10] S. Gleyzes, S. Kuhr, C. Guerlin, J. Bernu, S. Deleglise, U. Busk Hoff, M. Brune, J-M. Raimond, S. Haroche, "Quantum jumps of light recording birth and death of a photon in a cavity," Nature, vol.446, pp. 297-300, 2007.

[11] I. Reznikoff, "On a realistic and discrete approach to physical time," HAL open archives, hal-03191083, v1, 2021.

[12] G. Berkeley, Essay towards a New Theory of Vision, contents: 43 and 117, 1709.

Related works:

[13] J. Levine, "Materialism and qualia: the explanatory gap," Pacific Philosophical Review, 64, pp. 354-361, 1983.

[14] B. A. Wallace, The taboo of Subjectivity: Toward a New Science of Consciousness, Oxford: Oxford, 2000.

[15] B. L. Lancaster, "On the relationship between cognitive models and spiritual maps," Journal of Consciousness Studies, 7 (11-12), pp. 231-250, 2000.

[16] J. Levine, Purple Haze: The Puzzle of Consciousness, Oxford: Oxford, 2001.

[17] J. McFadden, "Synchronous firing and its influence on the brain's electromagnetic field," Journal of Consciousness Studies, 9, pp. 23-50, 2002.

[18] I. V. Danilov, "Primary Data Entry", COGNITIVE 2021 (this volume).

Selective Information-Driven Learning for Producing Interpretable Internal Representations in Multi-Layered Neural Networks

Ryotaro Kamimura
Kumamoto Drone Technology and Development Foundation
Techno Research Park, Techno Lab 203
1155-12 Tabaru Shimomashiki-Gun Kumamoto 861-2202
and IT Education Center, Tokai Univerisity
4-1-1 Kitakaname, Hiratsuka, Kanagawa 259-1292, Japan
email: ryotarokami@gmail.com

***Abstract*—This paper aims to propose a new type of information-theoretic method to control information content stored and transmitted in neural networks. To make the meaning of information concretely interpretable, we introduce the selective information and a method to control it, called "selective information-driven learning". The new method is more suited for modeling neural learning than conventional information-theoretic measures, such as mutual information, because we can easily maximize and minimize the information, and we can interpret the meaning of information more concretely. The method was applied to the well-known wine data set. The experimental results show that the selective information could be maximized and minimized, and we could easily interpret the meaning of information in terms of the number of strong weights. In addition, the partial compression of multi-layered neural networks revealed that maximum information tended to be focused on output information, while minimum information tried to consider input information in addition to output information. Finally, collective weights, averaged over all compressed weights obtained in learning, were similar to the original correlation coefficients between inputs and targets, meaning that the selective information can disentangle complicated connection weights into simple, linear, and independent ones to be easily interpreted.**

Keywords-Selective information; mutual information; partial compression; collective interpretation; correlation coefficient

I. Introduction

Due to the requirement of right of explanation [1], many attempts have been made to explain how neural networks try to produce outputs. Because the inner mechanism of neural networks is far from being clarified, the black-box property has been taken into granted, in particular, in practical applications such as medical ones [2]. In those applications, the black-box has been considered to be not so critical as has been expected, because even the human body is a kind of black-box. The most important thing for application lies in the usability and prediction performance of adopted models. Though in neural networks, as well as machine learning, there have been many different types of methods to explain the network behaviors, they seem to suppose implicitly or overly this type of black box model. For example, in the field of convolutional neural networks, the majority of methods for interpretation, have been focused on the explanation of network behaviors with implicitly supposed black-boxed models. Since it is impossible to clarify the inner mechanism at the present stage, all we can do is to check how outputs can be changed in accordance with the inputs, namely, external explanation. The well-known methods such as the activation maximization, surrogate functions, local perturbations, layer-wise relevance propagation, and so on [3]–[7], seem to detect features, corresponding to the specific inputs. Though they try to make the maximally informative explanation [8], they seem to suppose implicitly the black boxed properties of inference mechanism. Those methods have been very effective in practical applications, in particular, in the cases of image data sets, because it is easy to understand intuitively the meaning of features detected by those methods. However, even though those methods with strong visual power can show how some important features can be detected in multiple layers in neural networks, it is far from understanding the inner mechanism of our intelligence [9].

In addition, when we try to deal with data sets whose meaning cannot be easily understood such as business data sets, more interpretable models to make the black-box whiter are needed [9], [10]. Even in the seemingly interpretable image data sets, the well-known Clever Hans phenomena [11] cannot be easily explained without understanding the overall context in the prediction. Naturally, there have been also some attempts to interpret the main and inner mechanism of human nervous systems, as well as human cognitive processes, dating back to the so-called "connectionism" approch to human cognition [12]–[14]. However, those attempts could not necessarily clarify the inner structure by which complicated cognitive processes can be explained [15].

Parallel to this connectionism approach to human cognition, there were important attempts to interpret the inner structure from the information-theoretic points of view. Linsker's maximum information preservation had an impact to the studies to understand and explain visual information processing, as well as human information processing in general [16]–[19]. Linsker's approach was a trigger to produce many different types of information-theoretic methods in neural networks [20]–[26]. Though the attempts may be promising in exploring the general information processing properties behind neural networks, the complexity of information measures, such as mutual information in computation have prevented those information-theoretic methods from being widely used in neural networks. In addition, we have another problem of difficulty in understanding the meaning of final internal repre-

 ISBN: 978-1-61208-847-1

sentations. Ironically, by introducing the information-theoretic methods, the inner structure has become uninterpretable due to the abstract properties of those information-theoretic measures. Thus, we can say that the abstract property of information measures, such as mutual information made the information-theoretic methods themselves black-boxed, though they tried to understand and explain the inner mechanism of human information-theoretic processing.

In this context, the present paper tries to make the concept of information as concrete as possible and as interpretable as possible in terms of components of neural networks. We suppose that the information content can be described in terms of selectivity of components of neural networks. When the selectivity of components increases, they tend to have more information content. The research on the selectivity have been well discussed in neuro-sciences [27]–[33]. In addition, in the interpretation methods in the field of convolutional neural networks, many interpretation methods have actually focused on the detection of selectivity of some parts of neural networks to the coming inputs [34]–[39].

We here represent information content stored in neural networks in terms of selectivity of components. When the selectivity increases, more information can be stored. Thus, the information dealt with in this paper is called "selective information", and a learning method by using this selective information should be called "selective information-driven learning". Because we can maximize completely this concrete selective information and at the same time minimize it, we can interpret states with maximum and minimum information by which we can infer the actual internal representations with intermediate information content.

The paper is organized as follows. In Section 2, we first define the selective information and how to increase and decrease this selective information in the name of selective-information-driven learning. Then, we briefly explain how to compress connection weights partially and step by step to examine the information flow in multiple layers of neural networks. In Section 3, we present the results on the well-known application to the classification of wines. We first explain that the selective information could be maximized and minimized, producing different connection weights. When the selective information increases, more individually separated weights appeared, while more collective behaviors of several neurons appeared when the selective information decreased. Finally, we show that generalization and selectivity-based interpretation may be contradictory to each other, but the contradiction can be solved by supposing two types of information for different levels.

II. Theory and Computational Methods

We here explain the concept of selective information and how to compute it for multi-layered neural networks. In addition, we present how to compress partially multi-layered neural networks to examine the outputs from hidden layers.

A. Selective Information-Driven Learning

Now, let us begin with the definition of selectivity and selective information. For simplicity's sake, we compute the selectivity between the second and third layer denoted by the notation (2,3) in Figure 1. For the first approximation, we suppose that the strength of connection weights can be obtained by their absolute values. When the strength of absolute values of weights are larger, neurons connected with these weights are more strongly connected. As shown in Figure 1(a), in an initial stage of learning, all connection weights are randomly connected in all layers. When, the selective information is maximized, only one connection weight is connected with the corresponding neuron. When the selective information is minimized, we can have different types of states with minimum information. For example, as shown in Figure 1(c), all connection weights have equal and strong absolute values. Stronger connection weights may cause troubles in improving generalization and interpretation, we decrease the strength of connection weights as much as possible as shown in Figure 1(d). In this paper, we try to decrease the strength of connection weights when we try to minimize the selective information.

Then, we can obtain the absolute values of original weights from the second to the third layer,

$$u_{jk}^{(2,3)} = \mid w_{jk}^{(2,3)} \mid \tag{1}$$

where the notation (2,3) denotes the transition from the second to the third layer. Then, we normalize these values by their maximum one, which can be computed by

$$g_{jk}^{(2,3)} = \frac{u_{jk}^{(2,3)}}{\max_{j'k'} u_{j'k'}^{(2,3)}} \tag{2}$$

where maximum operation is over all connection weights between two layers.

Now, by using this normalized strength, selective information for the second to the third layer (2,3), can be computed by

$$G^{(2,3)} = n_2 n_3 - \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} g_{jk}^{(2,3)} \tag{3}$$

where n_2 and n_3 denote the number of neurons in the second and the third layer. Then, we try to increase or decrease this selective information, and in particular, we actively control this information to produce appropriate internal representations. When only one connection weight has some strength, while all the others are zero, the selective information is maximized $(n_2 n_3 - 1)$. On the contrary, when all connection weights have the same strength, the selective information become zero. In the extreme case, when no connection weights exist between two layers, the selective information is minimized by definition, because all connection weights have the same strength of zero.

To maximize the selective information, we must control the normalize strength g_{jk}. However, when we try to decrease this selective information, we need to reduce the strength

 ISBN: 978-1-61208-847-1

for all neurons between the layers. For this, we introduce an complement case of the original normalized strength

$$\bar{g}_{jk}^{(2,3)} = 1 - g_{jk}^{(2,3)} \tag{4}$$

This means that, when the strength increases, this inverse one decreases. This inverse equation have an effect to reduce the strength of larger connection weights. When the strength becomes larger, the corresponding connection weights are forced to be smaller. Then, by combining those two, we have a unified one

$$h_{jk}^{(2,3)} = \alpha g_{jk}^{(2,3)} + \bar{\alpha} \bar{g}_{jk}^{(2,3)} \tag{5}$$

where the parameter α ranges between zero and one. Then, the connection weights at the $t+1$th learning step is simply computed by

$$w_{jk}^{(2,3)}(t+1) = h_{jk}^{(2,3)} w_{jk}^{(2,3)}(t) \tag{6}$$

When the parameter α increases, weights tend to be more affected by the force to increase the strength. On the contrary, when the α decreases, the weights are more strongly affected by the force to decrease the strength as shown in Figure 1(d). When the information is maximized, only one connection weight is used to connect one with the other, while in the minimum state, all neurons are equally connected with each other. Usually, the learning starts with randomly initialized weights, corresponding to an intermediate information state. From this intermediate state, we can increase and decrease information flexibly, and we interpret the meaning of information in terms of the number of strong weights. On the contrary, information, for example, even when mutual information can be defined for the connection weights, it is extremely difficult to understand the actual meaning in terms of components of neural networks. However, the selective information, though defined very simply, can represent the content of mutual information concretely in terms of the number of strong connection weights.

1) Partial Compression : To examine information flow in multiple hidden layers, we here use the partial compression in which a multi-layered neural network is compressed gradually up to the simplest one without hidden layers in Figure 2.

Let us show how to compress a multi-layered neural network gradually from the input to the output layer. For this, we suppose that the number of neurons in all the hidden layer is the same. This does not necessarily exclude the variable number of neurons. Actually, the number of neurons can be reduced by the effect of selective information maximization, where the number of strong connection weights can be increased or decreased.

The first partial compression in Figure 2(b) lies in connecting the input and output layer

$$w_{ir}^{(1,2,7)} = \sum_{q=1}^{n_6} w_{iq}^{(1,2)} w_{qr}^{(6,7)} \tag{7}$$

where the notation (1,2,7) represents compressed weights up to the second layer, and n_6 is the number of neurons in the sixth

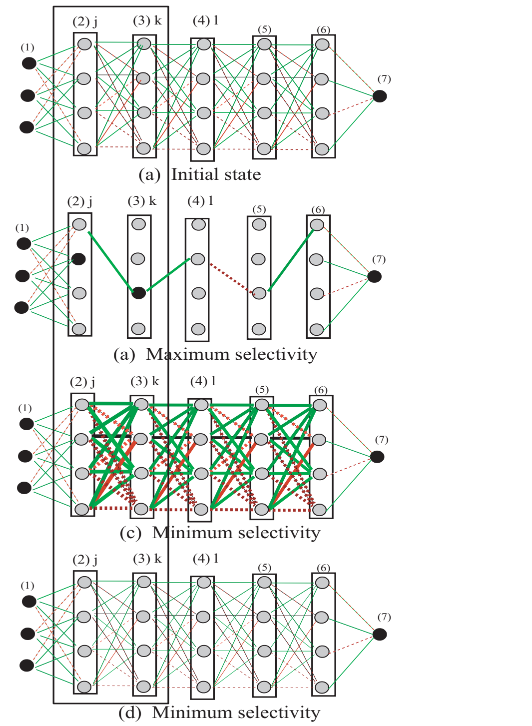


Fig. 1. Network architecture with seven layers, including five hidden layers (a) for active information maximization (b), information minimization No.1 (c) and selective informaiton minimization No.2 (d).

layer. The second partial compression in Figure 2(c) begins with connecting the first two connection weights

$$w_{ik}^{(1,2,3)} = \sum_{j=1}^{n_2} w_{ij}^{(1,2)} w_{jk}^{(2,3)} \tag{8}$$

where the notation (1,2,3) denotes compression up to the second layer. Then, by combing it with weights to the output layer, we have the second partial compression

$$w_{ir}^{(1,3,7)} = \sum_{q=1}^{n_6} w_{iq}^{(1,3)} w_{qr}^{(6,7)} \tag{9}$$

where the notation (1,3,7) shows the compression up to the third layer with the weights to the output layer. In the same way, we gradually combing the remaining connection weights, and finally, we can compress all connection weights

$$w_{ir}^{(1,6,7)} = \sum_{q=1}^{n_6} w_{iq}^{(1,5,6)} w_{qr}^{(6,7)} \tag{10}$$

where the notation (1,6,7) denote compression up to the sixth layer with the seventh layer or output layer, and (1,5,6) denotes compression up to the fifth layer.

III. Results and Discussion

We applied the method to the well-known wine data set to show how well the method could maximize and minimize selective information, and we could easily interpret the meaning of connection weights. In addition, the method could

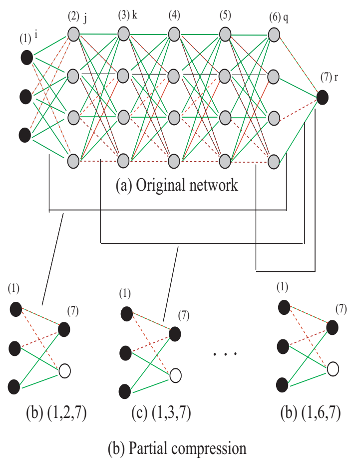


Fig. 2. Network architecture with seven layers, including five hidden layers (a) to be compressed gradually and partially into the simplest ones (b).

produce weights, which was close to the original correlations between inputs and targets. This means that the method could disentangle complicated relations into the simple ones.

A. Experimental Outline

The experiments aimed to predict red and white wine from the north of Portugal [40], based on twelve input variables. The number of patterns was 4898 white and 1599 red wines. As shown in Figure 3, the number of hidden layers was ten with ten neurons for each layer. We compressed multi-layered neural networks partially (b) and fully (c) for interpreting the information flow in multiple layers. The parameter α decreased from 1 to 0, which means that the selective information could be maximized and minimized. The number of learning steps was 100, and within each learning step, there were several sub-steps to assimilate the effect of selective information, ranging from 5 to 10. Though we can completely maximize and minimize the selective information, the selective information minimization was accompanied by the weight strength reduction, making the learning impossible when the information becomes closer to a minimum state. Thus, we made the effect of selective information minimization weaker for the stable learning. We should note that without considering the stability of learning, we could completely minimize the selective information. We used the scikit-learn neural network package with all default setting except the number of learning steps (epochs) and tangent-hyperbolic activation function, and naturally, connection weights were modified by the composite function to control the selective information. Those default values were used to make the reproduction of the present results as easy as possible.

B. Selective Information and Mutual Information

First, we try to show that the selective information can be controlled flexibly, reducing the strength of connection weights. In addition, the selective information can be easily understood in terms of the number of weights, while mutual information cannot give concrete meaning to the final results.

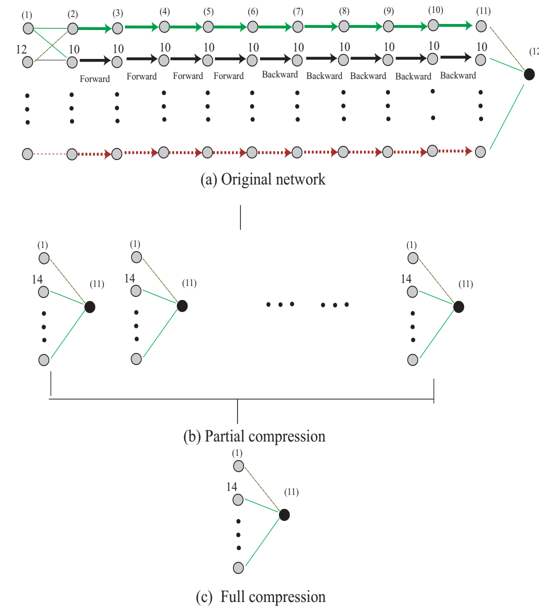


Fig. 3. Network architecture with 12 layers (10 hidden layers) in which each hidden layer has 10 neurons, a series of partial compression (b) and full compression (c) for the wine data set.

Figure 4 shows the selective information (left), mutual information (middle) and averaged absolute weight strength (right), when the parameter α decreased from 1(a) to 0(e). As shown in the leftmost figure of Figure 4(a), the selective information increased rapidly and close to its maximum value (10 by 10 by definition) in the end. Then, when the parameter decreased from 0.7(b) to 0(e), the selective information decreased gradually. When the parameter was zero in the leftmost figure of Figure 4(k), the selective information became slightly larger than that obtained when the parameter was 0.2 in Figure 4(j). As mentioned, to stabilize the learning, we reduced the effect of selective information when the parameter decreased and in particular, close to zero. Without this constraint on the selective information minimization, it could be reduced to almost an minimum point of zero. Thus, when the parameter decreased, the selective information was forced to be smaller as can be expected. In the same way, the figures in the middle shows mutual information, where mutual information increased immediately up to almost its maximum value. Then, when the parameter decreased, mutual information decreased gradually, and final close to zero. Then, the rightmost figures show the average strength of absolute weights, which were forced to be smaller by decreasing the parameter α.

As above mentioned, we can easily interpret the values of selective information. When the selective information is higher, the number of stronger weights becomes smaller. On the contrary, when the selective information is smaller, the number of stronger weights connecting with specific neurons becomes smaller, and all the weights become equal in their strength. At the beginning of learning, the random initial states are usually given, and the selective information in this case should be close to the middle of 50, meaning that the

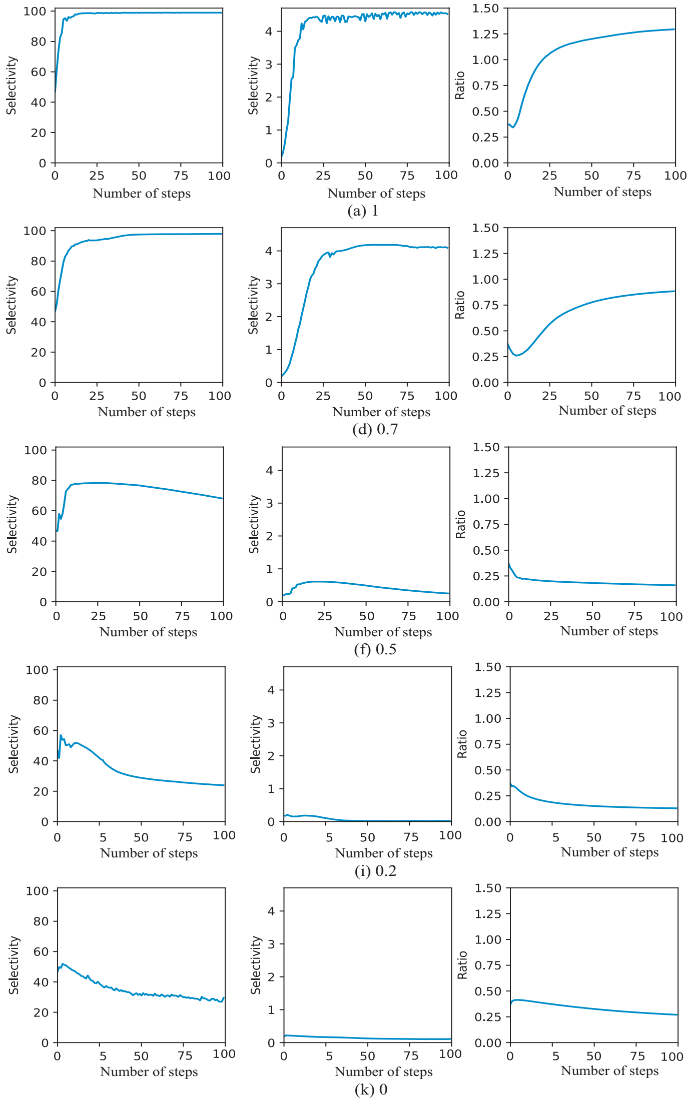


Fig. 4. Selective information (left), mutual information (middle) and averaged weights (right) when the parameter α decreased from 1(a) to 0(k) for the wine data set.

fifty percent of all connection weights should have stronger weights, though the strength of weights are different from each other. Then, when the selective information increases, the number of stronger weights becomes smaller and smaller and in the end when the maximum information state is reached, where only one connection weight becomes strong, while all the others are zero. On the contrary, when the selective information becomes smaller, the number of strong weights becomes smaller, and all weights become equally small. In the extreme case, all connection weights becomes zero, which is also a minimum selective information state by definition.

Then, we should examine how connection weights changed when the parameter also changed. Figure 5(a) shows connection weights when the parameter was one, namely, when only selective information maximization was applied. As can be expected, only a small number of connection weights among many became stronger, while all the others became very small. When the parameter decreased to 0.7 in Figure 5(b), a neuron connected with many neurons appear, which could be seen over weights close to the input and output layer. When the parameter was 0.5 in Figure 5(c), we could have an explicit pattern that all neurons in the precedent layers tended to be connected with many neurons in the subsequent layers, and vice versa. When the parameter was decreased to 0.2 in Figure 5(d), this tendency was further enhanced, and all neurons were explicitly connected with all the other neurons. Finally, when the parameter was set to zero, and only selective information minimization was applied, the tendency became slightly weaker. This can be explained by the fact that we made the effect of selective information weaker for the stability of learning. These results show that when the selective information increases, individual connection weights tend to behave independently of other weights. Then, when the selective information becomes smaller, connection weights behave collectively, and a neuron tend to be connected with many other neurons.

C. Partial Compression

Then, we tried to examine how information from the inputs and outputs were transmitted in multiple layers. The results show that the selective information maximization tried to capture output information mainly, while information minimization tried to take into account input information in addition to output information.

Figure 6 shows partially compressed weights when the parameter decreased from 1 (a) to 0 (e). When the parameter was one in Figure 6(a), and the selective information is maximized, partial compressed weights in the intermediate layers, were very weak in their strength, and only in the final compression state, namely, in the full compression, the weights became stronger. This means that information on the outputs should be necessary to form the compressed weights. When the parameter decreased from 0.7 (b) to 0.2 (d), gradually, in the initial stages of partial compression, located on the left hand side, connection weights became relatively stronger. This means that when the selective information becomes smaller, information on inputs was taken into account. Finally, when the parameter became zero, namely, when only the selectivity minimization was applied, the states of partial compression became similar to those by the selective information maximization only used in Figure 6(a).

For more clearly presenting this tendency, we computed the standard deviation of absolute connection weights of partially compressed weights in Figure 7. For example, Figure 7(a) shows the standard deviation of compressed weights when the parameter was one, and only selective information maximization was used. As can be seen in the figure, the standard deviation of weights from the first compression to the ninth compression was very small. When only all weights were compressed fully, the standard deviation became larger, meaning that the final connection weights played important roles to form the appropriate compressed weights. Then, when the parameter decreased from 1 (a) to 0.1 (j), gradually,

 ISBN: 978-1-61208-847-1

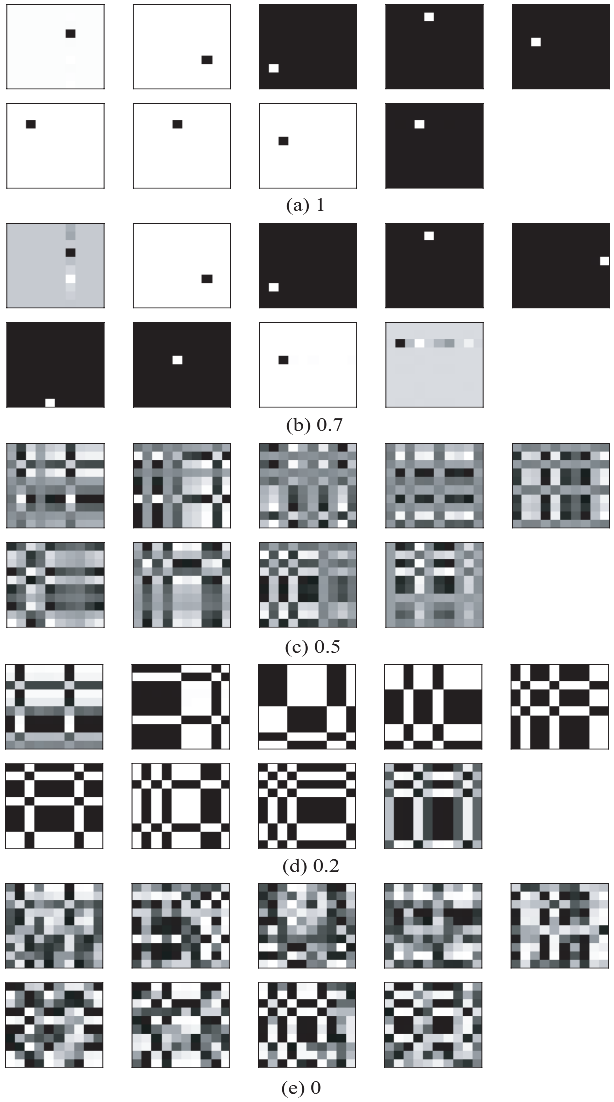


Fig. 5. Weights of hidden layers when the parameter α decreased from 1 (a) to 0 (e) for the wine data set.

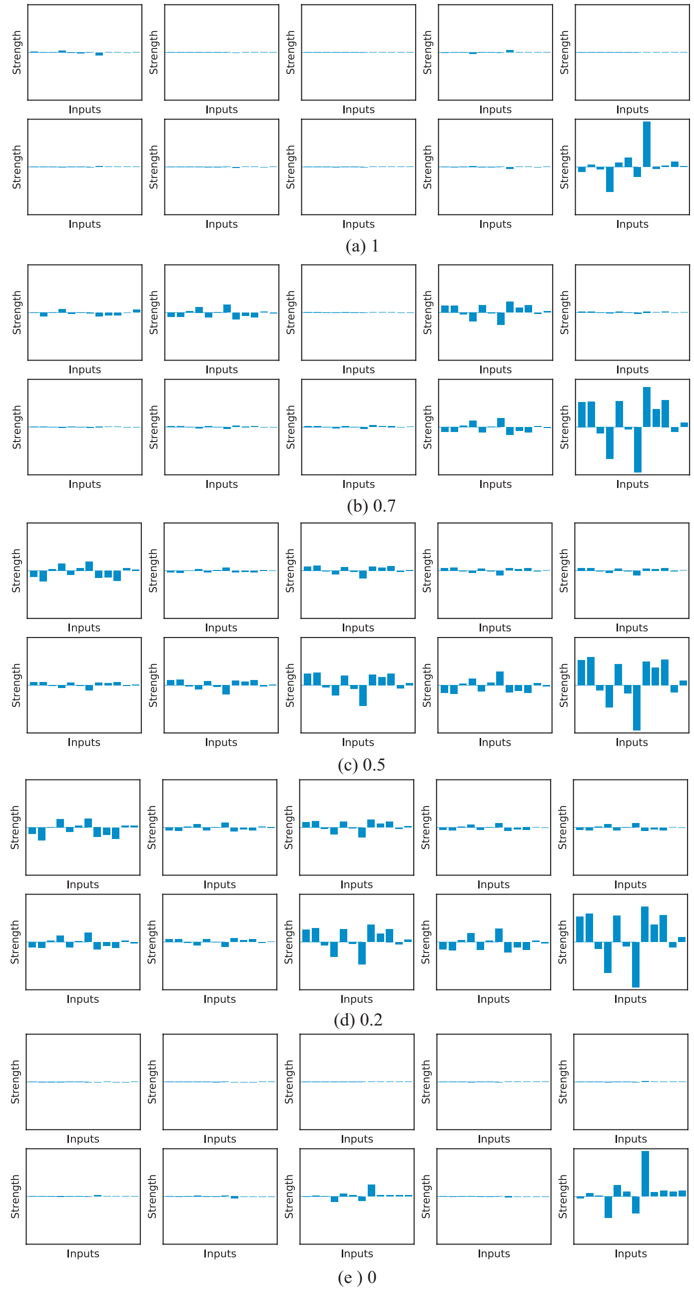


Fig. 6. Partially compressed weights when the parameter α changed from 1(a) to 0(e).

the standard deviation of absolute weights of initial stages of partial compression became stronger, though those compressed weights were still weaker. This means that when the selective information is decreased, connection weights in the intermediate layers tended to have some information, probably, on inputs. Finally, when the parameter was zero in Figure 7(k), the partially compressed weights became similar to those obtained when the parameter was one. This cannot be easily interpreted, but we infer that input information acquired in the intermediate layers, can be obtained only when selective information maximization and minimization effect are combined with each other. Or, as mentioned, we made the effect of selective information weaker, which may be the main cause of this state with the zero parameter.

D. Compressed Weight Comparison

We here compare the compressed weights with those by the other conventional methods. The results show that the final compressed weights were quite similar to the original correlation coefficients between inputs and targets. This means that the selective information has an effect to disentangle connection weights to have simpler, linear and independent relations between inpts and outputs.

Figure 8(a) shows the correlation coefficients between inputs and targets of the original data set. When the parameter was one in Figure 8(b), the correlation between the original correlations in Figure 8(a) and compressed weights was 0.673. When the parameter decreased to 0.7, 0.5, 0.2, the correlation increased to 0.953, 0.946 and 0.958, which were almost perfect correlations. Those correlations were higher than 0.937 of

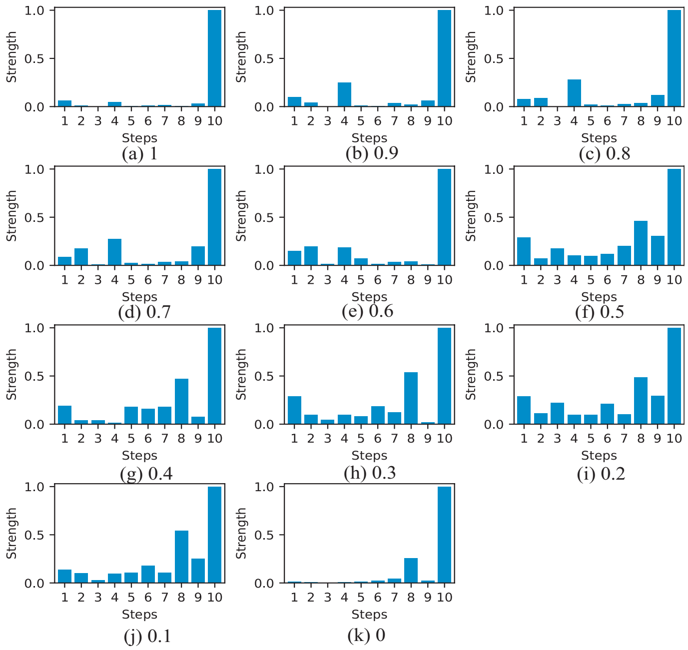


Fig. 7. Standard deviation of partially compressed weights when the parameter α decreased from 1 (a) to 0 (k) by 0.1 for the wine data set.

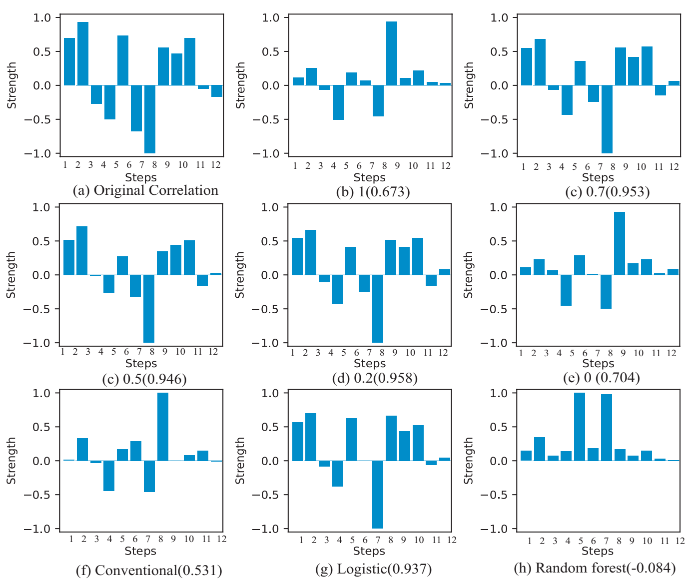


Fig. 8. Correlation coefficients (a), compressed weights for α=1 down to 0 (b)-(e), compressed weights by the conventional method (f), and the regression coefficients by the logistic regression analysis (g) and prediction importance by the random forest method (h) for the wine data set.

conventional logistic regression analysis in Figure 8(g). By the conventional method without selective information, the correlation was only 0.531 in Figure 8(f), and in addition, the random forest produced the worst correlation of -0.084 in Figure 8(h).

Finally, we examined relations between correlations and generalization. The results show that the interpretation and generalization were inversely correlated. Thus, we need to make an attempt to unify improve interpretation and generalization. Figure 9(a) shows the correlation coefficients between

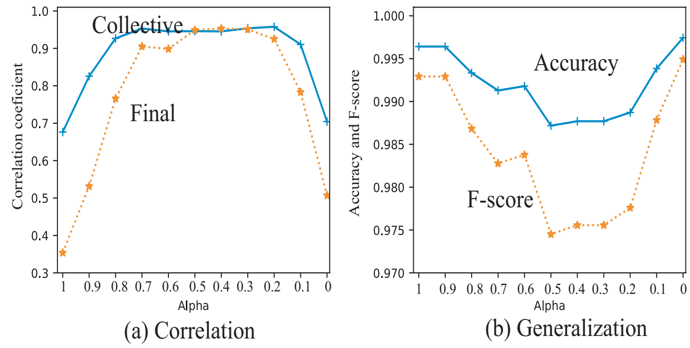


Fig. 9. Correlation coefficients (a) and generalization accuracy (b) for the wine dat set.

the correlations of original data set and compressed (final) and collective weights (collective). The compressed weights were ones, obtained when the learning steps was the final one, namely, 100 step in this case. On the other hand, the collective weights were obtained by averaging all compressed weights for all intermediate learning steps. As shown in the figure, the collective weights produced always higher correlations than the compressed weights. This means that the simple average of all compressed weights in learning can increase the correlation between the original correlation and compressed weights. Figure 9(b) shows generalization accuracy and F-measure, where the accuracy was always larger than the F-measure. Comparing Figure 9(a) and (b), we can conclude that the correlations were inversely related to generalization performance, though decrease in generalization was considerably small. This can be easily interpreted by using the selectivity of neurons and connection weights. When the selectivity of components of neural networks increases, and they tend to respond to the inputs very specifically, they cannot deal with less specific inputs naturally. Thus, we make a compromise between selectivity and generalization or interpretation and generalization. At the first glance, it seems to be impossible to make this kind of compromise between them, but this type of contradiction can be easily be solved by supposing selective information maximization and minimization operating in two different contexts or levels. We should explore this possibility of contradiction resolution as a future study of this paper.

IV. Conclusion

The present paper aimed to propose a new type of information-theoretic method called "selective information-driven learning". The selective information is introduced to measure the selectivity of components in neural networks to replace conventional mutual information, because it can easily be interpreted in terms of the number of strong weights, while conventional mutual information can not be easily interpreted in terms of components of neural networks. The new method was applied to the well known wine data set. The experimental results showed that the selective information could be maximized and at the same time minimized within the same framework. The interpretation can be possible in terms of number of strong connection weights, which is easier to understand, compared with the conventional mutual information. In addition, by partially compressing multi-layered neural networks, we could find that the selective information maximization is focused on output information, while the selective information minimization tried to detect input information as well.

The results confirmed that the explicit interpretation of internal representations is possible by the present method. However, it was observed that better interpretation is not necessarily followed by improved generalization. This is because the selectivity for improve interpretation may be harmful to improved interpretation, needing an ability to responding well non-specific and ambiguous cases. Thus, we need to make further studies on unifying improved interpretation and generalization.

REFERENCES

[1] B. Goodman and S. Flaxman, “European union regulations on algorithmic decision-making and a right to explanation,” *arXiv preprint arXiv:1606.08813*, 2016.

[2] M. Sendak, M. C. Elish, M. Gao, J. Futoma, W. Ratliff, M. Nichols, A. Bedoya, S. Balu, and C. O’Brien, “” the human body is a black box” supporting clinical decision-making with deep learning,” in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 99–109, 2020.

[3] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in *European conference on computer vision*, pp. 818–833, Springer, 2014.

[4] R. C. Fong and A. Vedaldi, “Interpretable explanations of black boxes by meaningful perturbation,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3429–3437, 2017.

[5] L. M. Zintgraf, T. S. Cohen, T. Adel, and M. Welling, “Visualizing deep neural network decisions: Prediction difference analysis,” *arXiv preprint arXiv:1702.04595*, 2017.

[6] M. Ancona, E. Ceolini, C. Öztireli, and M. Gross, “Towards better understanding of gradient-based attribution methods for deep neural networks,” *arXiv preprint arXiv:1711.06104*, 2017.

[7] G. Montavon, A. Binder, S. Lapuschkin, W. Samek, and K.-R. Müller, “Layer-wise relevance propagation: an overview,” in *Explainable AI: interpreting, explaining and visualizing deep learning*, pp. 193–209, Springer, 2019.

[8] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, *et al.*, “Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai,” *Information Fusion*, vol. 58, pp. 82–115, 2020.

[9] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, 2019.

[10] M. N. Angenent, A. P. Barata, and F. W. Takes, “Large-scale machine learning for business sector prediction,” in *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, pp. 1143–1146, 2020.

[11] S. Lapuschkin, S. Wäldchen, A. Binder, G. Montavon, W. Samek, and K.-R. Müller, “Unmasking clever hans predictors and assessing what machines really learn,” *Nature communications*, vol. 10, no. 1, pp. 1–8, 2019.

[12] D. Rumelhart and J. M. et al., *Parallel Distributed Processing*, vol. 1. MA: MIT Press, 1986.

[13] D. E. Rumelhart and D. Zipser, “Feature discovery by competitive learning,” *Cognitive Science*, vol. 9, pp. 75–112, 1985.

[14] D. E. Rumelhart, G. E. Hinton, and R. Williams, “Learning internal representations by error propagation,” in *Parallel Distributed Processing* (D. E. Rumelhart and G. E. H. et al., eds.), vol. 1, pp. 318–362, Cambridge: MIT Press, 1986.

[15] D. E. Rumelhart and J. L. McClelland, “On learning the past tenses of English verbs,” in *Parallel Distributed Processing* (D. E. Rumelhart, G. E. Hinton, and R. J. Williams, eds.), vol. 2, pp. 216–271, Cambrige: MIT Press, 1986.

[16] R. Linsker, “Self-organization in a perceptual network,” *Computer*, vol. 21, no. 3, pp. 105–117, 1988.

[17] R. Linsker, “How to generate ordered maps by maximizing the mutual information between input and output signals,” *Neural computation*, vol. 1, no. 3, pp. 402–411, 1989.

[18] R. Linsker, “Local synaptic learning rules suffice to maximize mutual information in a linear network,” *Neural Computation*, vol. 4, no. 5, pp. 691–702, 1992.

[19] R. Linsker, “Improved local learning rule for information maximization and related applications,” *Neural networks*, vol. 18, no. 3, pp. 261–265, 2005.

[20] S. Becker, “Mutual information maximization: models of cortical self-organization,” *Network: Computation in Neural Systems*, vol. 7, pp. 7–31, 1996.

[21] K. Torkkola, “Nonlinear feature transform using maximum mutual information,” in *Proceedings of International Joint Conference on Neural Networks*, pp. 2756–2761, 2001.

[22] K. Torkkola, “Feature extraction by non-parametric mutual information maximization,” *Journal of Machine Learning Research*, vol. 3, pp. 1415–1438, 2003.

[23] J. M. Leiva-Murillo and A. Artés-Rodríguez, “Maximization of mutual information for supervised linear feature extraction,” *Neural Networks, IEEE Transactions on*, vol. 18, no. 5, pp. 1433–1441, 2007.

[24] M. M. Van Hulle, “The formation of topographic maps that maximize the average mutual information of the output responses to noiseless input signals,” *Neural Computation*, vol. 9, no. 3, pp. 595–606, 1997.

[25] J. C. Principe, D. Xu, and J. Fisher, “Information theoretic learning,” *Unsupervised adaptive filtering*, vol. 1, pp. 265–319, 2000.

[26] J. C. Principe, *Information theoretic learning: Renyi’s entropy and kernel perspectives*. Springer Science & Business Media, 2010.

[27] E. L. Bienenstock, L. N. Cooper, and P. W. Munro, “Theory for the development of neuron selectivity,” *Journal of Neuroscience*, vol. 2, pp. 32–48, 1982.

[28] A. Schoups, R. Vogels, N. Qian, and G. Orban, “Practising orientation identification improves orientation coding in v1 neurons,” *Nature*, vol. 412, no. 6846, pp. 549–553, 2001.

[29] L. E. White, D. M. Coppola, and D. Fitzpatrick, “The contribution of sensory experience to the maturation of orientation selectivity in ferret visual cortex,” *Nature*, vol. 411, no. 6841, pp. 1049–1052, 2001.

[30] H. Ko, S. B. Hofer, B. Pichler, K. A. Buchanan, P. J. Sjöström, and T. D. Mrsic-Flogel, “Functional specificity of local synaptic connections in neocortical networks,” *Nature*, vol. 473, no. 7345, pp. 87–91, 2011.

[31] J. F. Jehee, S. Ling, J. D. Swisher, R. S. van Bergen, and F. Tong, “Perceptual learning selectively refines orientation representations in early visual cortex,” *Journal of Neuroscience*, vol. 32, no. 47, pp. 16747–16753, 2012.

[32] M. V. Peelen and P. Downing, “Category selectivity in human visual cortex,” 2020.

[33] B. J. Bongers, A. P. IJzerman, and G. J. Van Westen, “Proteochemometrics–recent developments in bioactivity and selectivity modeling,” *Drug Discovery Today: Technologies*, 2020.

[34] A. S. Morcos, D. G. Barrett, N. C. Rabinowitz, and M. Botvinick, “On the importance of single directions for generalization,” *arXiv preprint arXiv:1803.06959*, 2018.

[35] I. Rafegas, M. Vanrell, L. A. Alexandre, and G. Arias, “Understanding trained cnns by indexing neuron selectivity,” *Pattern Recognition Letters*, vol. 136, pp. 318–325, 2020.

[36] J. Ukita, “Causal importance of low-level feature selectivity for generalization in image recognition,” *Neural Networks*, vol. 125, pp. 185–193, 2020.

[37] M. L. Leavitt and A. Morcos, “Selectivity considered harmful: evaluating the causal impact of class selectivity in dnns,” *arXiv preprint arXiv:2003.01262*, 2020.

[38] W. J. Johnston, S. E. Palmer, and D. J. Freedman, “Nonlinear mixed selectivity supports reliable neural computation,” *PLoS computational biology*, vol. 16, no. 2, p. e1007544, 2020.

[39] M. L. Leavitt and A. S. Morcos, “On the relationship between class selectivity, dimensionality, and robustness,” *arXiv preprint arXiv:2007.04440*, 2020.

[40] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis, “Modeling wine preferences by data mining from physicochemical properties,” *Decision Support Systems*, vol. 47, no. 4, pp. 547–553, 2009.

The Role of Resonance in the Development and Propagation of Memes

Muneo Kitajima*, Makoto Toyota† and Jérôme Dinet‡
* Nagaoka University of Technology, 1603-1 Kamitomioka Nagaoka, Niigata, Japan, Email: mkitajima@kjs.nagaokaut.ac.jp
† T-Method, Chiba, Japan, Email: pubmtoyota@me.com
‡ Université de Lorraine, 91 avenue de la Libération, 54000 Nancy, France, Email: jerome.dinet@univ-lorraine.fr

Abstract*—Meme is a type of behavior that is passed from one member of a group to another, not through the genes but by other means. This paper claims that the concept of resonance should play an important role in the propagation of memes among group members and the development of meme in individual members. Dinet et al. pointed out that the concept of resonance originally issued from physics that has been successfully applied to cognitive processes for behavior selection; the Model Human Processor with Real-Time Constraints (MHP/RT) model proposed by Kitajima and Toyota is a unique model that incorporates a resonance mechanism to connect perceptual-cognitive-motor (PCM) processes that work synchronously, and memory processes that work asynchronously with the environment. This paper elaborates the resonance mechanism implemented in MHP/RT and its associated multi-dimensional memory frames from three perspectives: how resonance works in a single-action selection process; what types of memes can exist in multi-dimensional memory frames; how PCM processes and memory develop from birth to the end of adolescence in terms of the detailed workings of resonance. This paper concludes with the implications of the mechanistic understanding of the role of resonance, the effect of resonance malfunction and the effect of the existence of memory that does not resonate with the real environment.

Keywords–Resonance; Meme; MHP/RT; Development.

I. Introduction

At the 0-th order approximation, a person interacts with the environment by running an endless stream of perceptions involving the external and internal environments through the five senses, i.e., taste, sight, touch, smell, and sound, via sensory neurons as parallel processing, and acting in response to the external environment using body parts, e.g., limbs, eye balls, and so on, via motor neurons using serial processing (see Figure 1). As s/he perceives the results of movement of his/her body parts as well as the changes in the external environment with the passage of time, the next cycle of perceptual-motor processes occur. Interneurons between the sensory neurons and motor neurons convert the input patterns to the output patterns – this constitutes a Perceptual-Cognitive-Motor (PCM) process.

Starting from this basic cycle (see Figure 1), Kitajima and Toyota [3][4] constructed a comprehensive theory of action selection and memory, Model-Human Processor with Real-time Constraints (MHP/RT), that provides a basis for constructing any models for an acting person (see Figure 2). MHP/RT is an extension of Model Human Processor proposed by Card, Moran, and Newell [5] that can simulate routine goal-directed behaviors. The processes involved in action selection is a dynamic interaction that evolves in the irreversible time dimension. The purpose of MHP/RT is to explain the following three facts: 1) the fundamental processing mechanism of the brain is Parallel Distributed Processing (PDP) [6], which is referred to as Organic PDP (O-PDP) system in the development of MHP/RT; 2) human behavior emerges as a result of competition of the dual processes of System 1, fast *unconscious* processes for intuitive reaction with feedforward control that connect perception with motor movements, and System 2, slow *conscious* processes for deliberate reasoning with feedback control; this is called Two Minds [7]; 3) human behavior is organized under happiness goals [8].

MHP/RT consists of two parts. The first part comprises

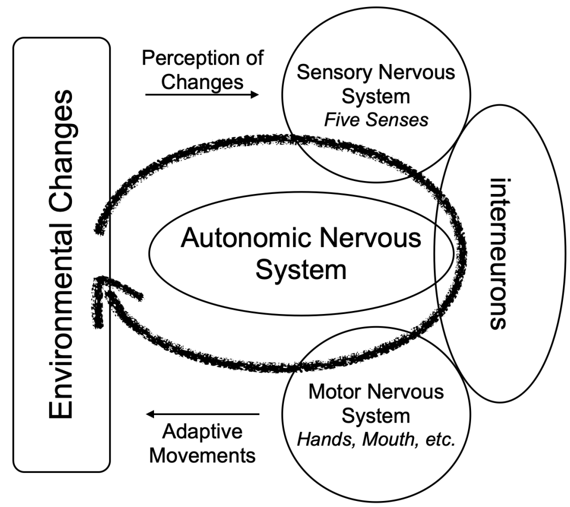
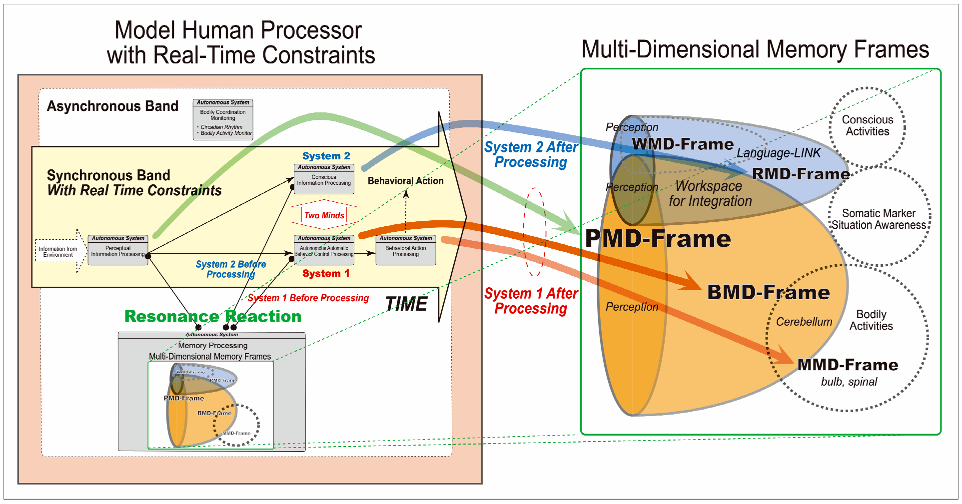


Figure 1. Continuous cyclic loop of perception and movement ([9, Figure 1]).

TABLE I. Newell's time scale of human action. (Adapted from Newell [10, page 122, Fig. 3-3]).

Scale (sec)	Time Units	System	World (Theory)
10^7	months		
10^6	weeks		SOCIAL
10^5	days		BAND
10^4	hours	Task	
10^3	10min	Task	RATIONAL
10^2	minutes	Task	BAND
10^1	10sec	Unit Task	
10^0	1sec	Operations	COGNITIVE
10^{-1}	100ms	Deliberate Act	BAND
10^{-2}	10ms	Neural Circuit	
10^{-3}	1ms	Neuron	BIOLOGICAL
10^{-4}	100μsec	Organelle	BAND



WMD (Word MD)-frame is the memory structure for language. It is constructed on a very simple one-dimensional array.
RMD (Relation MD)-frame is the memory structure associated with the conscious information processing. It combines a set of BMD-frames into a manipulable unit.
BMD (Behavior MD)-frame is the memory structure associated with the autonomous automatic behavior control processing. It combines a set of MMD-frames into a manipulable unit.
PMD (Perceptual MD)-frame constitutes perceptual memory as a relational matrix structure. It incrementally grows as it creates memory from the input information and matches it against the past memory in parallel.
MMD (Motion MD)-frame constitutes behavioral memory as a matrix structure. It gathers a variety of perceptual information as well to connect muscles with nerves using spinals as a reflection point. In accordance with one's physical growth, it widens the range of activities the behavioral action processing can cover autonomously.

Figure 2. MHP/RT and the distributed memory system implemented as multi-dimensional memory frames (modified from [4, Figure 3]).

cyclic PCM processes (see Figure 2, left), in which PDP for these processes is implemented in hierarchically organized bands with characteristic times for operations by associating relative times (not absolute) to the PCM processes that carry out a series of events *in synchronous with* changes in the external environment. Table I shows the bands, i.e., biological, cognitive, rational, and social bands, as defined by Newell [10]. It should be noted that there is a gap between two adjacent bands; these two bands are non-linearly connected and therefore it is inappropriate to understand the phenomena that happen across these bands by constructing a linear model. The phenomena occur by connecting what happens in a band to what happens in its adjacent band non-linearly. A mechanism is required to connect the phenomena; MHP/RT suggests that this connection is provided by the resonance mechanism.

The bottom left and middle portions of Figure 2 show the autonomous memory system consisting of multi-dimensional memory frames of perception, motion, behavior, relation, and word. These memory frames store information associated with the corresponding autonomous processes defined in the PCM processes. The middle portion of Figure 2 shows the five memory frames and their relationship with the PCM processes. The right portion of Figure 2 provides brief explanations of the respective memory frames. The important feature of the memory system is that it works *asynchronously* with the external environment. MHP/RT assumes that the *synchronous* PCM processes, including the perceptual system, System 1, System 2, and the motor system, and the asynchronous memory system communicate with each other through a resonance mechanism. The concept of resonance has been borrowed from physics to describe the linkage between the asynchronous memory system and the synchronous PCM processes. As Dinet et al. [1] suggested, apprehension of psychological phenomena using concepts issued from physics is useful because the majority of the interactions, including psychological interactions, between humans and the environment (social or physical environment) can be derived from physical processes.

The PCM process cycle in a human being continues from his or her birth until death in the ecological system that consists of the self and the environment. The linkage between the memory system and the PCM processes through resonance supports the development of the PCM processes. At the same time, the PCM processes are accompanied by changes in the connections of neurons that constitute the multi-dimensional memory frames. The purpose of this paper is to understand the development process from the viewpoint of the memory-PCM linkage through resonance.

This paper is organized as follows. Section II focuses on the role of resonance in the use of memory and its relationship with the changes in memory when a human being interacts with the environment. Section III extends the snap-shot view of resonance explored in the previous section in the time dimension to understand the nature of external objects that generate sustainable resonance. We define these sustainable objects as "memes" and this section describes how memes propagate over time and space. Section IV describes the development of the PCM processes and the multi-dimensional memory frames from birth to the end of adolescence. Section V concludes the paper by summarizing the contents and pointing out implications derived from the considerations of the role of resonance described in this paper.

II. Memorizing the Results of Resonance

The network of sensory neurons, motor neurons, and interneurons shown in Figure 1 develop as a human being interacts with the external environment in time. This section starts by describing a general feature of the development process in Section II-A, followed by the descriptions of what is memorized in Section II-B, and how the resonance mechanism relates to memory usage and memory formation in Section II-C.

A. Forming an O-PDP System

Life emerges in a preexisting structure, which is the global environment that surrounds the human being who is born, and exists as a system that operates under the mechanism of PDP.

 ISBN: 978-1-61208-847-1

Life has evolved through adaptation to the global environments, which are characterized by the following preexisting structures:

- **Atmosphere:** fluctuates at the meteorological scales showing chaotic behavior,
- **Sea:** has tidal currents and the surface layer is affected by the atmosphere showing chaotic behavior,
- **Periodic circular structures:** stable periods of rotation and revolution around the sun with a close-circular orbit as a planet of the solar system,
- **Gravity:** upper vs. lower directions, defined by the earth's gravity, and
- **Energy:** future vs. past directions, defined by the direction of energy flow as the dissipative structure of the earth.

Life is formed under these preexisting structures and the direction of the evolution of life is determined by the pressure caused by these structures. Under the dissipative structural space, i.e., the earth, the fundamental structural pressure is a prerequisite to life, which is born and active in it, and prescribes the direction of life evolution. Life is formed as an adaptive body with the functional and structural features that work most efficiently in the environment having the features best conceived by the four concepts of Goals, Operators, Methods, and Selection rules (GOMS) [5].

In the global environment before life occurred, there existed multiple-layered structures equipped with means of communicating information in a multi-dimensional space, including such physical and chemical entities as light, sound, heat, ion, etc. Under these preexisting conditions, life emerged. Life interacts with the multi-dimensional environment by using perception and motion as interfaces to it. Assuming the size of dimension of perception and that of motor to be M and N, respectively, the function of interface is represented as mappings in the $M \otimes N$ space. Specific numbers of M and N reflect the preexisting conditions of the environment.

B. *Interneurons to memorize effective $M \otimes N$ mappings*

Table I shows that human actions are hierarchically organized in four bands. In general, in-band closed processes are executed in feedforward. Processes carried out in the upper bands provide feedback to the processes carried out in the lower bands. Life activities perform mappings of M-dimensional perceptual input (Biological band) to N-dimensional motor output (Biological band) by the system of interneurons that belongs to the Cognitive band. Activities at an upper band emerge, each of which is associated with a part of the entire sequence of activities, i.e., a sub-sequence, performed at a lower band. The action sequence belonging to the lower band is segmented into sub-sequences, each of which corresponds to one of activities in the upper band; due to the fluctuations in processing times of the upper-band activities, relativization of processing times and functionalization of sub-sequences occur. It is possible to establish an association between the processing in the upper band activity (Cognitive band) with an effective sequence of $M \otimes N$ mappings in the lower band (Biological band) to form a hard-wired circuit of interneurons that is characteristic of this processing, which is the process of memory formation and development. Once the memory is formed, life activities perform mappings of M-dimensional perceptual input to N-dimensional motor output by using the memory stored in the system of interneurons.

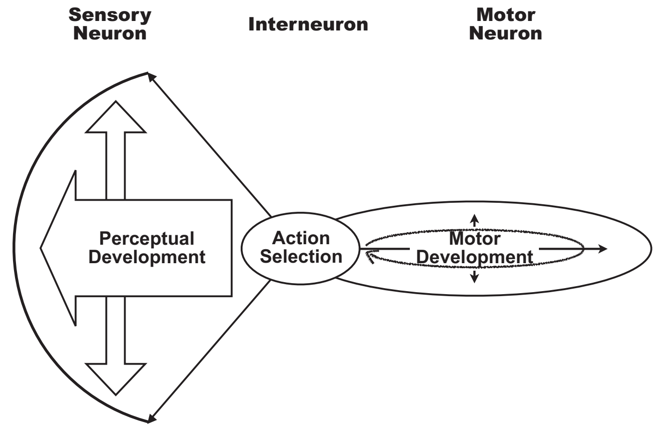


Figure 3. Development of the sensory nervous system and the motor nervous system, and interneurons connecting them with action selection process. ([11, Figure 1]).

In addition, trajectories of the life activities fluctuate, which provides opportunities to the O-PDP system to develop by adaptation.

The genealogy of DNA of vertebrates suggests that Perception, Interneuron, and Motion as the basis for development. Figure 3 shows that PIM is the basis of the formation of the concept of the body based on neural circuits and the whole is formed as PIM develops. Perception (P) captures various kinds of environmental changes via the sensors with different properties (*M*-dimension). Motor movement (M) is carried out continuously and cyclically from one's birth with gradual development via the accuracy and strength (*N*-dimension). Interneurons (I) memorize effective interlocked relationships between P and M to form neural circuits naturally in the form of feedforward (System 1) and more complex feedback to increase the effectiveness of reactions (System 2). PIM develops by expanding behavioral-ecological bandwidth. Behavioral ecological categories of vertebrates such as acquiring food, raising children, and so on, are almost identical and within a limited range. Everyday human life is performed cyclically in a behavioral ecological band, which is expanded by performing new actions to realize a goal within the limited behavioral ecological categories and adding these to the existing band.

Actions are realized as an *ad hoc* adaptive combination of various elements of the O-PDP system each time. Quasi-stable relationships between elements are generated by local relativization of the time relationship between the participating elements and added to the existing band. Since each process of performing actions is autonomous, it is possible for multiple processes to be initiated in a certain environment, resulting in *coincidental* parallel activities. Even if the timing of the execution of two or more parallel processes changes, it is possible for all of the processes to be completed within a certain time range with or without any relationships in the results. If the overall results are good, behavior selection in the future might be changed by this memory associated with rewards. Arbitrarily activated processes through conscious System 2 thinking, which may or may not be directly related to the current on-going external situation, can become a part of these expanded memories. Conscious thinking facilitates the expansion of memory.

 ISBN: 978-1-61208-847-1

C. Conscious/Unconscious Processes Before/After an Event

MHP/RT maps $M \otimes N$ with the help of memory (Figure 2, left). It assumes the multi-dimensional memory frames and the resonance mechanism for incorporating memory, i.e., interneurons, in the mapping process. Perception with the M-dimensional perceptual information resonates with the contents in the PMD-frame. The activation in the PMD-frame spreads to the other memory frames. The activated portions of memory frames resonate with the System 1 and System 2 processes and are included in these processes. In the left portion of Figure 2, the resonance processes are shown graphically by •——•. It is suggested that the mirror-neuron system is the physical/neurological support for the resonance in human behavior [1] that can be regarded as the process of $M \otimes N$ mapping. The mirror system is thought to directly match visual input from an observed action with a stored motor program for the same behavior (e.g. [12][13][14]). If that motor program is then executed, the result is imitation.

MHP/RT assumes that actions are carried out by feedforward processes (System 1) or feedback processes (System 2) with the help of the resonance mechanism for utilizing existing memory. The processing principle of MHP/RT is that it should work in one of four different modes when one looks at it from *a particular event* that occurred at the absolute time T [15]. These modes are called Four-Process in the MHP/RT architecture. In Figure 2, these four modes are indicated where appropriate to show how they are supported by the resonance mechanism to utilize the contents stored in memory (System 1 and System 2 Before modes), and how they affect in forming memory (System 1 and System 2 After modes) as described below.

1) Rise of Resonance: Two modes work before the event. In the time range of $T-\beta \leq t < T-\beta'$, MHP/RT uses memory for *consciously* preparing for what would happen in the future (*System 2 Before Mode*), and in the time range of $T - \beta' \leq t < T$, it *unconsciously* coordinates motor activities to the interacting environment (*System 1 Before Mode*), where $\beta' \sim$ 500msec and β ranges from a few seconds to hours, and even to months. In these two modes, the part of memory activated through resonance in response to perceptual processing could resonate with System 1 processing and System 2 processing (see Figure 2, left).

2) Convergence of Resonance: The other two modes work after the event. In the time range of $T < t \leq T+\alpha'$, MHP/RT *unconsciously* tunes the connections between sensory inputs and motor outputs for better performance for the same event in the future (*System 1 After Mode*), and in the time range of $T+\alpha' < t \leq T+\alpha$, it *consciously* recognizes what has happened and then modifies memory concerning the event (*System 2 After Mode*), where $\alpha' \sim$ 500msec and α ranges from a few seconds to minutes, and even to hours. In these two modes, the results of action selection for the event at T would be reflected in the network connections of the respective multi-dimensional memory frames (see Figure 2, middle).

III. Meme Propagation by Means of Resonance

Memory is represented as a network of neurons. It is initially activated through an M-dimensional multimodal sensory input through a resonance mechanism, a large part of which originates from the other human beings or artifacts they have created; then the activation spreads in the connected memory regions. The information generated by human beings propagates among them by means of imitative behavior in which resonance connects memory and the PCM processes to carry out imitation. This raises the question of what causes resonance to be maintained in humans. This section introduces the Structured Meme Theory [16] which provides an answer to this question.

A. What Can Resonate

The O-PDP system that defines a human being interacts with the environment, which consists of not only inanimate objects but also other O-PDP systems (people who interact with him/her). Resonance is the mechanism by which action is taken at appropriate times. Resonance increases when the organism is in System 1 and 2 Before Modes of the MHP/RT, but converges when in the System 1 and 2 After Modes which effectively update the internal neural networks for future resonance. This section describes what resonates in an O-PDP system. The resonance mechanism, which is implemented by means of mirror neurons [1] in the human brain, is incorporated in the MHP/RT as the mechanism for making available the part of memory activated by perception in System 1 and 2 (see Figure 2, left).

The cyclic loop of perception and movement shown in Figure 1 carries out continuous updates of network connections of constituent neurons during the convergence of resonance period in the MHP/RT. Human behavior is structured in four bands (biological, cognitive, rational, and social bands) [10], and memory is constructed as five distinctive multi-dimensional memory frames (see Figure 2, middle). This suggests that the information is coming from an environment where the informant could either be other humans or inanimate objects. This information will generate resonance in the recipient of the information in a specific way that reflects the states of the development of the multi-dimensional memory frames. For a set of given external stimuli, a human being generates resonance by using the current memory for carrying out the next action and updates it. On the one hand, the PCM processes utilize the resonance mechanism to create the subsequent actions. On the other hand, the external entities, that make resonance generate internally, are created ultimately by humans. Human action associated with a band of the Newell's time scale could be associated with the part of memory that participates via the resonance mechanism.

From the above argument, it is plausible to assume that a human could create an entity that makes the others generate resonance for performing the subsequent actions and vice versa. This assumption represents what was meant by the term "meme" that Richard Dawkins [17] coined in the 1970s, which was conceptual and was not defined clearly. The Structured Meme Theory (SMT) proposed by Toyota et al. [16] defines memes as entities that represent the information associated with the objects that the brain can recognize.

Figure 4 depicts how memes propagate in the reality field. The process of propagation is facilitated by symbolization. A symbolized meme enables human beings to think at abstract levels. Suppose that Object-O_1, Object-O_2, or Object-O_3 appear in the environment. Each of these resonates with memory by using patterns stored in the PMD-frame and causes activation of RMD-, BMD-, and MMD-frames. The part of memory that is activated, Pattern-P_a (see Figure 4, bottom), will be used in action selection through resonance. Any entities in the environment that match Pattern-P_a are treated in the

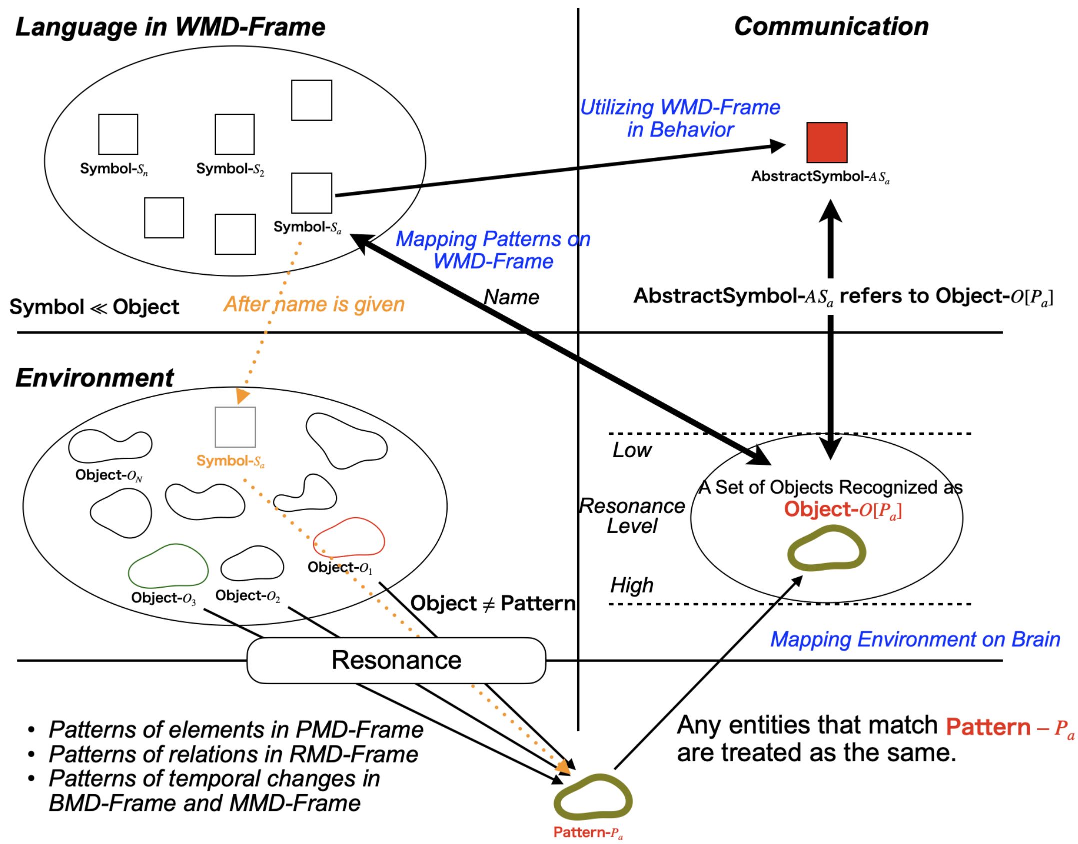


Figure 4. Propagation of Meme

same way in the action selection process as Object-$O[P_a]$. Pattern-P_a will be given a unique name Symbol-S_a and it will be stored in WMD-frame. In the future, when Symbol-S_a appears in the environment in someone's utterances, it might resonate with Pattern-P_a. Symbol-S_a could be associated with Object-O_1, Object-O_2, or Object-O_3. Therefore Symbol-S_a effectively functions as its abstraction, AbstractSymbol-AS_a, when it is used in communication. It is reasonable to assume that the higher the resonance level of Symbol-S_a becomes, the longer AbstractSymbol-AS_a is maintained and inherited as an effective communication medium.

B. Structured Meme

The recent consensus is that the range of information inherited by the genes is limited to physical functions and infantile behavior. Human beings need to acquire basic behavioral skills and communication skills through the experience of acting in the environment. As described in Section III-A, memes that can cause resonance constitute an important part of experience. This section explains that memes are structured in such a way that the memory is structured in multi-dimensional memory frames.

Figure 5 expands Pattern-P_a, that appears at the bottom of Figure 4, and the processes shown in Figure 5 as "Mapping Patterns on Brain" and "Mapping Patterns on WMD-frame" in order to show which portions of the neural networks would participate while a human being utilizes the WMD-frame in communication. As shown by Figure 4, symbols in the WMD-frame are gradually incorporated into the environment in the form of a thesaurus, i.e., lists of words in groups of synonyms and related concepts, languages used for person-to-person communication (individual language), which might include not only direct but also metaphorical uses, and languages used in cultural contexts (cultural language), in which appropriate understanding of common sense that has been established in the specific community, is essential for successful communication. Thesauruses, individual languages, and cultural languages increase their complexity in this order in terms of the patterns they are linked with the objects in the environment. Thesauruses are associated with the objects in the environment that are encoded in the neural networks in the initial development stage from the birth to 3 years. Individual languages are associated with not only the objects in the environment but also the symbols that have already been incorporated in the environment. The same is true for cultural languages.

The process of "Mapping patterns on symbols in WMD-frame" shown in Figure 4 can be subdivided into three processes based on the degree of complexity of mapping. The patterns that were mapped on the thesauruses, individual languages, and cultural languages are shown as Action-Level Meme, Behavior-Level Meme, and Culture-Level Meme (see Figure 5), respectively, that were introduced in the Structured Meme Theory proposed by Toyota et al. [16] . The mechanism

 ISBN: 978-1-61208-847-1

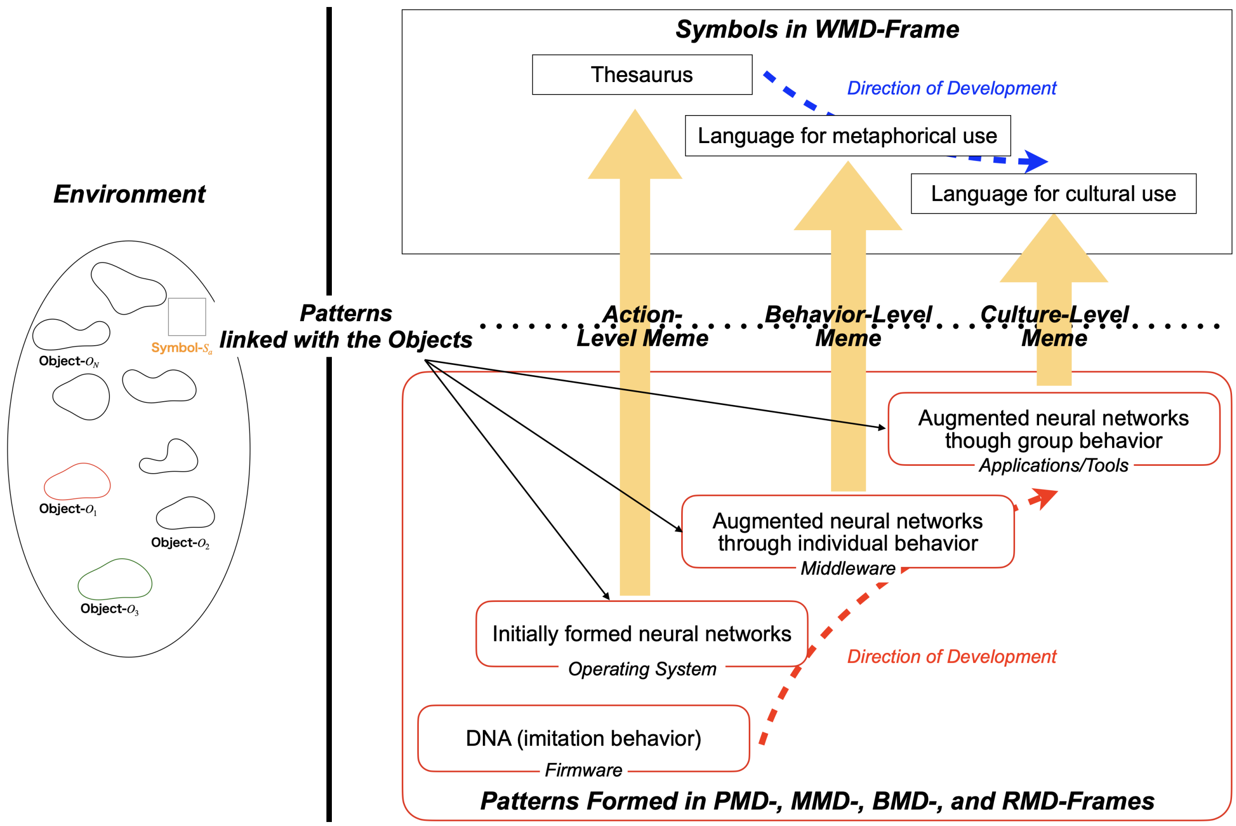


Figure 5. Structure of meme

by which the three levels of memes and genes inherit information is analogous to an information system. Genes serve as firmware that mimics behavior-level activities. Action-level memes serve as the operating system that defines general patterns of spatial-temporal behavioral functions. Behavior-level memes represent middleware that extends the general patterns to concrete patterns. Culture-level memes act as application tools that extend the concrete patterns to the ones that work in a number of groups of people.

The relationships between the three levels of memes and multi-dimensional memory frames are as follows:

- Action-level memes represent bodily actions stored in the MMD-frame;
- Behavior-level memes represent behaviors in the environment stored in the BMD-frame; and
- Culture-level memes represent culture stored in the RMD-frame and the WMD-frame.

IV. Development of PIM and Multi-dimensional Memory Frames

In Section II, we described the mechanism by which PIM develops with the Four-Process by utilizing the resonance mechanism. Section III described how the contents of multi-dimensional memory frames, which could resonate with the PCM processes running synchronously with the environment, are maintained among human beings through communication. This section describes the changes in the contents of PIM and multi-dimensional memory frames over time as human beings grow up from birth to the end of adolescent. This explains human development in terms of PIM and multi-dimensional memory frames (Figure 5, bottom).

This section elaborates the four stages in PIM development as described in [9]. Table II shows the age range, rate of autonomous development of perception and motion, and the method of connecting sensory and motor neurons via interneurons and the resultant control method. Figure 6 depicts schematically the development of PIM (the sensory nervous system and the motor nervous system characterized by the functioning of interneurons connecting them (Figure 3)) and the development of multi-dimensional memory frames that shows the structured inter-connections of interneurons along with each stage (Figure 2, center). The multi-dimensional memory frames that develop at a particular stage are shown by adding shades.

A. Stage 1

At $0 \sim 6$ years of age, feedforward loops are the dominant control mechanism and these establish fundamental relationships between MMD- and BMD-frames by means of “uplinking.” In the first half of this period, at $0 \sim 3$ years of age, sensory and motor neurons experience rapid autonomous development. Interneurons connect them by subordinate-intervention, resulting in feedforward control (Figure 6-1A). During this period, human beings establish inter-connections between PMD- and MMD-frames, and PMD-, MMD-, and BMD-frames. The former is related to action-level memes and the latter, to behavior-level memes. These make integrated movements of bodily actions possible on the basis of the relationships be-

Copyright (c) IARIA, 2021. ISBN: 978-1-61208-847-1

TABLE II. PIM DEVELOPMENT PROCESS

Stage	Age	Speed of Autonomous Development		Method for Establishing Connection	Resulting Control Mechanism
		Perception	Motion	Interneurons	
1	0∼ 3	Rapid	Rapid	subordinate	Feedforward control
2	4∼ 7	Fast	Fast	memory-mediated	Limited feedback control
3	7∼ 12	Stable	Stable	memory-mediated-proactive	Wide feedback control
4	13∼ 18	Settled	Settled	memory-mediated-autonomous	Extended feedback control

tween the input from the perceptual system in the PMD-frame and the output expressed as reflexive movements in the MMD- and BMD-frames, for example, simple utterances (Figure 6-1B).

B. Stage 2

In the latter half of this period, at $4 \sim 6$ years of age, there is rapid autonomous development of sensory and motor neurons. Interneurons connect these by memory-mediated interventions, resulting in feedback controls (Figure 6-2A). During this period, human beings acquire the skill of behaving in relation with other persons and the methods for conversing with others such as the ability to explain using simple syntax. Acquisition of utterances and combined motor movements become possible by using the RMD-frame that connects simple syntax of symbols in the WMD-frame by linking PMD-, BMD-, and MMD-frames. At this stage, the conscious processes of System 2 Before and System 2 After that make feedback controls possible are initiated (Figure 6-2B).

C. Stage 3

Later, at $7 \sim 12$ years of age, parallel distributed activities are almost complete. There is stable autonomous development of the sensory and motor neurons. Interneurons connect them by memory-mediated proactive interventions, resulting in wider feedback controls (Figure 6-3A). During this period, human beings acquire the skill of first-order logical thinking by using letters or symbols and that of cooperation with other persons. These activities facilitate the development of interconnections among the multi-dimensional memory frames, resulting in very complex networks. Figure 6-3B) shows the development of interconnections in the RMD- and WMD-frames. The key is the presence of symbols that intervene between the various input and output connections; these symbols are related to culture-level memes.

Structural development precedes procedural development. The former refers to the development of connections between PMD- and MMD-frames with the support of the BMD-frame. The BMD-frame stores repetitive sequences of motor actions in the MMD-frame as single units; therefore, the action sequences in the BMD-frame can be carried out unconsciously. The latter refers to the development of connections between WMD- and RMD-frames. The connections start from spoken informal language to more abstract structural representations of formal language. The WMD-frame stores sequences of symbols as language dissociated from the real world; the RMD-frame stores a set of BMD-frames into a manipulable unit. The connections between the contents in the WMD-frame and those in the RMD-frame give reality to the contents in the WMD-frame. Structural development and procedural development interact with each other in the RMD-frame occasionally and proceed in parallel. Procedural development in the RMD-frame is critical to extend the range of behavior. Once elements in the RMD-frame are compiled in the System 2 After mode, the compiled rule can be used to initiate a behavior with longer action steps.

D. Stage 4

Finally, at age $13 \sim 18$ years, parallel distributed activities are complete. There is settled autonomous development of sensory and motor neurons. Interneurons connect these using memory-mediated autonomous intervention, resulting in even wider feedback control (Figure 6-4A). During this period, feedback loops are established; these are used to form language processing circuits in a single-layer, WMD-frame. During this period, the ability of logical writing and the generation of grammatically correct sentences by using ordinary language affects significantly the development process. The WMD-frame can evolve autonomously and extensively; part of the WMD-frame is associated with the BMD-frame and ultimately connected with the real world but the rest of the WMD-frame is dissociated from the real world because there is no connection between the WMD-frame and RMD-frame (Figure 6-4B).

V. CONCLUSION

Resonance connects remote systems enabling them to work jointly. Human beings act in the environment by using two systems: a cyclic PCM system that works in synchronization with the environment and an autonomous memory system that operates asynchronously with the environment. Based on MHP/RT, this paper describes the hypothesis that the PCM system and the memory system interact with each other via a resonance mechanism. First, this paper shows that a single action selection carried out by the PCM system was conceived as a mapping of the input from the environment represented in M-dimensional information in the PMD-frame on the output to the environment represented in N-dimensional motor actions in the MMD-frame, connected with the BMD- and RMD-frames, and discusses the role played by resonance in the PCM processes defined by the MHP/RT and updates of the memory represented by the multi-dimensional memory frames. Second, in situations where the environment comprises objects created by humans, i.e., artifacts such as symbols, manners, language, and culture, this paper discusses how artifacts are maintained over time and space with the help of resonance. A necessary condition for the survival of artifacts is that they resonate with human beings. This paper introduces action-level, behavior-level, and culture-level memes to represent resonating entities stored in memory. Third, we also described the human development process from birth to the end of adolescence in terms of the changes in PIM and memory, taking into account the changes in memes that human beings develop in memory as they grow older.

The argument this paper has profound implications for cases where there is malfunction of resonance, and where memory that fails to resonate as reality develops.

An example of the first case is Autism Spectrum Disorder (ASD). Is ASD a resonance problem? Imitation is a potentially crucial aspect of social cognitive development [18] because imitation is an efficient tool for two main adaptive functions: learning and communication. Imitation-based communication is possible through the use of the two aspects of imitation: imitating and being imitated [19]. These two aspects give rise to two roles that partners can exchange by taking turns while they synchronize matched activities. Neuroimaging studies of interactive imitation have shown that such communicative systems involve a coordination of bottom-up and top-down processes. In accordance with some authors (e.g. [20]) and based on the MHP/RT model, we hypothesize that perceptive-motor resonance plays a more central role in imitation in infancy than does a rational evaluation of the observed action. Thus, because individuals with ASD have difficulties with imitation, the nature of which remains unclear, it is plausible to postulate that one of the main problems of ASD is related to difficulties with the mechanisms underlying resonance. Several studies have shown that very young children with ASD (M Age $<$ 48 months; for a synthesis: [21]) used imitation less often when copying the actions of others, spent less time looking at others' faces and more time looking at the actions of others; attentional, social and executive factors underlie different aspects of imitation difficulties in this population. In other words, ASD is characterized by a deficiency in the development of imitation capacity, and this deficiency implies a deficit in mapping neural codings for actions between sensory and motor modalities, rather than in motivation or executive function [22]. Thus, ASD could be characterized by abnormal development of these mappings and that of resonance mechanisms.

Another important implication of our hypothesis is related to the autonomous development of the WMD-frame at the later stage of development. This paper showed that part of the contents in the WMD-frame do not necessarily have reality, which is established by linking the WMD-frame with the chain of RMD-, BMD-, and MMD-frames. Systematic and logical knowledge as a group norm could become part of the WMD-frame through education. It could resonate with the symbols that refer to the knowledge but would not resonate with *real* objects in the environment. The larger the non-resonant part of memory, the less the person will resonate with reality in the environment.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 20H04290. The author would like to thank Editage (www.editage.com) for English language editing.

References

[1] J. Dinet and M. Kitajima, "The Concept of Resonance: From Physics to Cognitive Psychology," in COGNITIVE 2020 : The Twelfth International Conference on Advanced Cognitive Technologies and Applications, 2020, pp. 62–67.

[2] M. Kitajima and M. Toyota, "Simulating navigation behaviour based on the architecture model Model Human Processor with Real-Time Constraints (MHP/RT)," Behaviour & Information Technology, vol. 31, no. 1, 2012, pp. 41–58.

[3] M. Kitajima, Memory and Action Selection in Human-Machine Interaction. Wiley-ISTE, 2016.

[4] M. Kitajima and M. Toyota, "Decision-making and action selection in Two Minds: An analysis based on Model Human Processor with Realtime Constraints (MHP/RT)," Biologically Inspired Cognitive Architectures, vol. 5, 2013, pp. 82–93.

[5] S. K. Card, T. P. Moran, and A. Newell, The Psychology of Human-Computer Interaction. Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.

[6] J. L. McClelland and D. E. Rumelhart, Parallel Distributed Processing: Explorations in the Microstructure of Cognition : Psychological and Biological Models. The MIT Press, 6 1986.

[7] D. Kahneman, "A perspective on judgment and choice," American Psychologist, vol. 58, no. 9, 2003, pp. 697–720.

[8] D. Morris, The nature of happiness. London: Little Books Ltd., 2006.

[9] M. Kitajima and M. Toyota, "Hierarchical Structure of Human Action Selections – An Update of Newell's Time Scale of Human Action," in Procedia Computer Science, BICA 2014. 5th Annual International Conference on Biologically Inspired Cognitive Architectures, vol. 41, 2014, pp. 321–325.

[10] A. Newell, Unified Theories of Cognition (The William James Lectures, 1987). Cambridge, MA: Harvard University Press, 1990.

[11] M. Kitajima and M. Toyota, "Guidelines for designing artifacts for the dual-process," in Procedia Computer Science, BICA 2015. 6th Annual International Conference on Biologically Inspired Cognitive Architectures, vol. 71, 2015, pp. 62–67.

[12] J. Decety, T. Chaminade, J. Grèzes, and A. Meltzoff, "A pet exploration of the neural mechanisms involved in reciprocal imitation," NeuroImage, vol. 15, no. 1, 2002, pp. 265 – 272. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811901909383[retrieved:March,2021]

[13] M. Iacoboni and M. Dapretto, "The mirror neuron system and the consequences of its dysfunction," Nature Reviews Neuroscience, vol. 7, no. 12, 2006, pp. 942–951. [Online]. Available: https://doi.org/10.1038/nrn2024[retrieved:March,2021]

[14] M. Iacoboni et al., "Cortical mechanisms of human imitation," Science, vol. 286, no. 5449, 1999, pp. 2526–2528. [Online]. Available: https://doi.org/10.1126/science.286.5449.252[retrieved:March,2021]

[15] M. Kitajima and M. Toyota, "Four Processing Modes of *in situ* Human Behavior," in Biologically Inspired Cognitive Architectures 2011 - Proceedings of the Second Annual Meeting of the BICA Society, A. V. Samsonovich and K. R. Jóhannsdóttir, Eds. Amsterdam, The Netherlands: IOS Press, 2011, pp. 194–199.

[16] M. Toyota, M. Kitajima, and H. Shimada, "Structured Meme Theory: How is informational inheritance maintained?" in Proceedings of the 30th Annual Conference of the Cognitive Science Society, B. C. Love, K. McRae, and V. M. Sloutsky, Eds. Austin, TX: Cognitive Science Society, 2008, p. 2288.

[17] R. Dawkins, The Selfish Gene. New York City: Oxford University Press, 1976.

[18] H. J. Stewart, R. D. McIntosh, and J. H. Williams, "A specific deficit of imitation in autism spectrum disorder," Autism Res, vol. 6, no. 6, Dec 2013, pp. 522–530.

[19] J. Nadel, "Perception-action coupling and imitation in autism spectrum disorder," Dev Med Child Neurol, vol. 57 Suppl 2, Apr 2015, pp. 55–58.

[20] M. Paulus, S. Hunnius, M. Vissers, and H. Bekkering, "Imitation in infancy: Rational or motor resonance?" Child Development, vol. 82, no. 4, 2011, pp. 1047–1057. [Online]. Available: http://www.jstor.org/stable/41289825[retrieved:March,2021]

[21] G. Vivanti, D. Trembath, and C. Dissanayake, "Mechanisms of imitation impairment in autism spectrum disorder," J Abnorm Child Psychol, vol. 42, no. 8, Nov 2014, pp. 1395–1405.

[22] J. H. Williams, A. Whiten, and T. Singh, "A systematic review of action imitation in autistic spectrum disorder," J Autism Dev Disord, vol. 34, no. 3, Jun 2004, pp. 285–299.

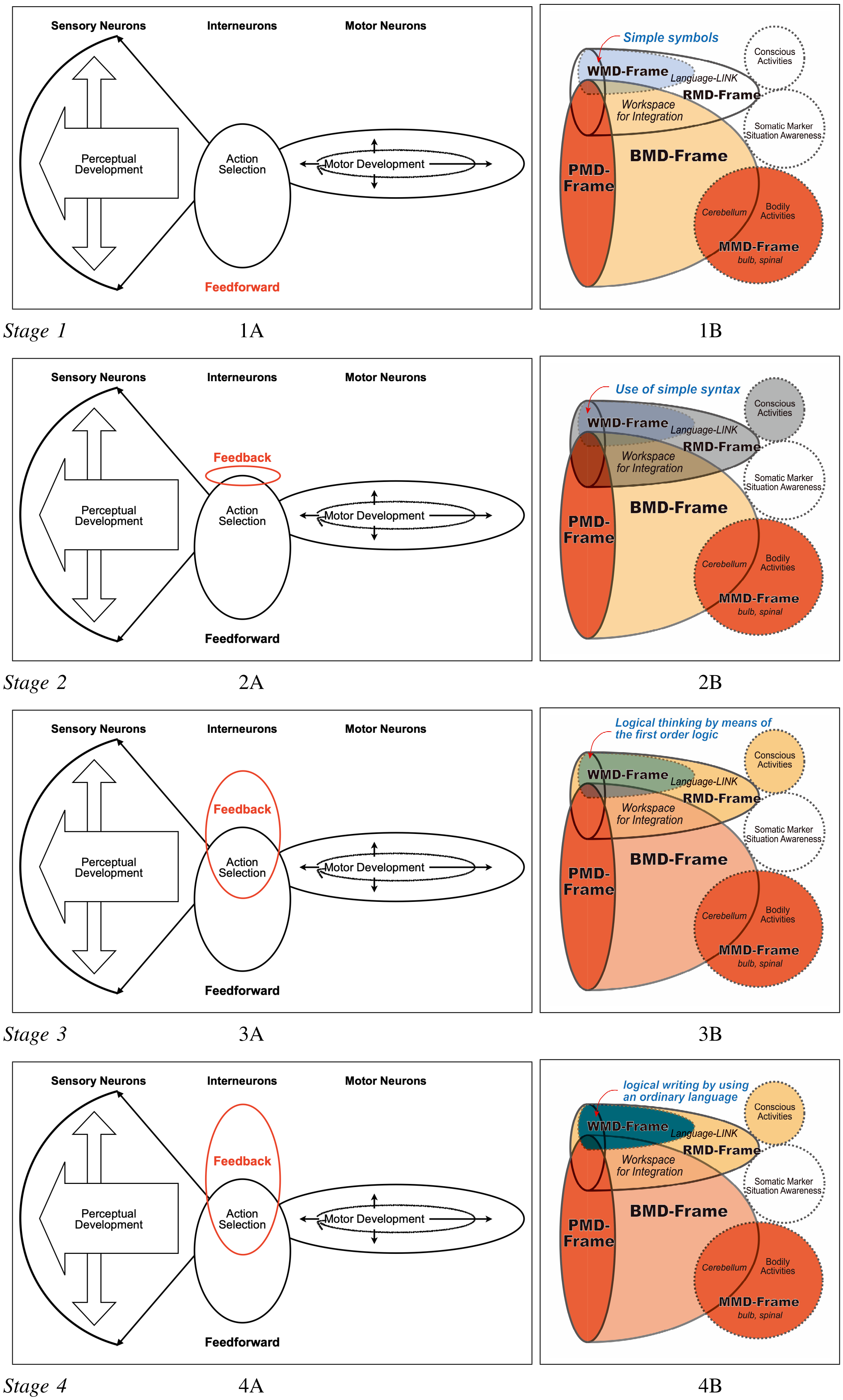


Figure 6. Development of PIM and multi-dimensional memory frames

Topic Modeling of StormFront Forum Posts

Grigorii Nazarko
Dept. of Business and Analytics
Arcada University of Applied Sciences
Helsinki, Finland
email: grinazarko (at) gmail.com

Richard Frank
School of Criminology
Simon Fraser University
Burnaby, Canada
email: rfrank (at) sfu.ca

Magnus Westerlund
Dept. of Business and Analytics
Arcada University of Applied Sciences
Helsinki, Finland
email: magnus.westerlund (at) arcada.fi

***Abstract*—Radicalized communities are actively using the internet to preach their ideas within society. Hence, it is crucial to research the content of such communities to understand their agenda and potentially take actions. The discussion is spread out in different forums on the Internet, but in this study, we used natural language processing technologies for revealing topics discussed in the oldest right-wing forum StormFront. As a result, we found the co-occurrence of discussed topics and real-world events, which means that the method can be used to track the agenda and changes in the community.**

Keywords-StormFront; topic model; information extraction; NLP; LDA

I. Introduction

Although most of the population uses the internet in a benign manner, there are groups, like political extremists of any conviction, that (ab)use the internet in order to facilitate the organization of violent events. In addition to simple coordination, extremists have been found to use the internet for all manner of benefits to their group, such as information provisioning, financing, networking, recruitment and information gathering [1]. Recently, there has also been evidence of some violent political extremists and terrorists being online immediately prior to their attacks [2], such as the Christchurch, New Zealand shooter in 2019 who, before live-streaming his attack on two mosques, posted his manifesto and shooting intentions on the imageboard "8chan" [3]. Further studies have also shown that, especially in extremist communities, online education, communication, and training are much more prevalent [4]. For example, of the extremists who plotted to attack government targets (as opposed to civilian targets) 83% "displayed traits of online learning", and in the context of violent extremism using Improvised Explosive Devices (IED), those who plotted to use the IEDs were 3.34 times more likely, while those who actually used IEDs were 4.57 times more likely, to use online sources for their information [4]. Clearly, extremists in general, make extensive use of the internet for their preparations.

As empirical research is suggesting that online resources are increasingly playing important roles in mobilizing extremists to violence, and that groups and movements mobilize themselves online, law enforcement agencies are increasingly recognizing the need to better understand these online mobilization efforts [5]. Online indicators of malicious behaviour are being developed by multiple law enforcement agencies (for an example, see [6]) to try to detect imminent dangers, by looking for specific actions by users, such as the posting of martyrdom videos or statements, attempting to mobilize others, or engaging in ideological discussions.

These indicators of imminent dangers are highly qualitative and difficult to detect. They are usually simple posts on online discussion forums or social media. One way to detect such dangers is to look for shifts in patterns in discussion. For example, a sudden shift from election to a discussion around the topic of "attacks" and "shooting" could be interesting. These sudden shifts in topics of discussion could indicate that the community is suddenly aware of a new trend or topic, which could be of interest to law enforcement. Thus, by monitoring online discussion forums, establishing baseline measures, and then comparing sudden changes in topics of interest, it should be possible to detect emerging threats. This is already in action in other domains, such as Google Trends [7] or Twitter Hashtag Trends [8] being used to model the emergence of news. In this paper, we evaluate the viability of using a similar approach using Natural Language Processing (NLP) to detect emerging trends within the online right-wing discussion forum StormFront with the eventual goal of creating a near-real-time monitoring system capable of identifying trends in discussion forums, as they emerge, giving law enforcement and security organizations the awareness capability to act as something is unfolding, rather than forensically after-the-fact.

In Section 2, we present a review of previous approaches to researching StormFront and its agenda. Section 3 introduces the method used to achieve the goal of the study. Section 4 provides a discussion on data exploration and data preparation. In Section 5 we review results of the study and finally in Section 6 we conclude the paper.

II. Literature review

Communities like StormFront can be analysed for the purpose of extracting information that help to describe them, to understand their motivations and progression. A traditional avenue of research is the manual gathering and processing of information. Dentice [9] performed a manual approach conducting interviews with representatives of different groups related to right-wing communities and by manual exploration of web resources, including the StormFront forum. The research provided an understanding of right-wing communities in Canada. Another manual approach analysed discussions from StormFront to reveal the understanding of the forum members' behaviour [10].

 ISBN: 978-1-61208-847-1

Manual approaches provide interesting and verifiable results, but the main drawback is the poor scalability in terms of high frequency data and the large volume of posts. Resource allocation provides a limit to manual research for analysing social media post en mass and may hinder the analysis of communities were the number of posts may be in the millions. One way to overcome this is to perform processing based on machine learning of the community data.

In one study, the effects of intergroup conflicts on StormFront and the behaviour of members were analyzed using NLP techniques (including sentiment analysis and word frequency) [11]. Another study described right-wing communities' behaviour using sentiment analysis on data from the StormFront forum [12]. Yet another approach used NLP and support vector machine to classify the StormFront posts depending on the rhetorical change in the posts [13]. However, the results of these studies provide a general description and understanding of right-wing communities; they do not provide the specifics of discussion topics and their temporal shift. We address this gap in our study of right-wing posts and detail a topic modeling approach and trend analysis for StromFront forum discussions.

III. Method

Topic modelling is an NLP task to discover abstract topics in a collection of documents. It assumes that each document in a set of documents (corpus) contains one or more topics that humans can understand; usually, those topics can be described by a set of interrelated keywords.

Latent Semantic Analysis [14] is a statistical approach that studies semantic distributions, relations between sets of documents and words that form concepts. The assumption is that closeness in meaning is related to closeness in textual distance. Latent Dirichlet Allocation (LDA) [15] is a topic modelling method, which is an extension of Probabilistic Latent Semantic Analysis (PLSA) [16]. LDA uses Dirichlet priors for the document-topic and word-topic distributions, which creates an improved generalisability compared to PLSA, i.e., it performs better on new data.

LDA and PLSA are both based on the assumption that each document in a corpus is a mixture over K topics, each topic, in turn, is a mixture over all the words from the dictionary.

LDA is a generative model. Generative models work by first determining the input and output distributions and then uses these distributions to generate additional synthetic input data. Then, the algorithm determines if the generated data represents the given data with an acceptable level of confidence. If the confidence level is not acceptable, then the model generates new data. This process runs in a loop.

In our case, the model generates bag-of-words distribution using topics per document distribution Θ and words per topic distribution φ. Θ and φ, in turn, are generated from a Dirichlet distribution with the parameters α and β.

Figure 1 shows the graphical representation of this process [15], where Z is topic-word and W word-document identities for a particular word in a document. M and N are a number of documents and a number of words in a particular document, respectively.

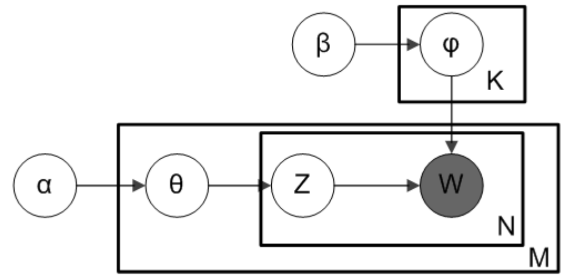


Figure 1. Graphical representation (plate notation) for LDA [15].

The next step is the estimation, whether the generative process has generated data close enough to the given data. There are various ways of performing this task. The LDA implementation, which we used, does it with the variational Bayes algorithm.

LDA takes a "bag of words" text representation as an input, which is a drawback of LDA because a large vocabulary grows the dimensionality, and as a result the computational performance of the method. LDA takes at minimum one hyper-parameter, the number of topics. The optimal number of topics can be determined with perplexity as a metric. Perplexity is a measurement of how good a probability distribution predicts a sample.

IV. Data

The goal of this paper is to identify emerging trends within online right-wing communities to detect possible new events and/or problems. A number of different right-wing online forums exist online, and for this paper a specific right-wing site, the online community StormFront, was chosen.

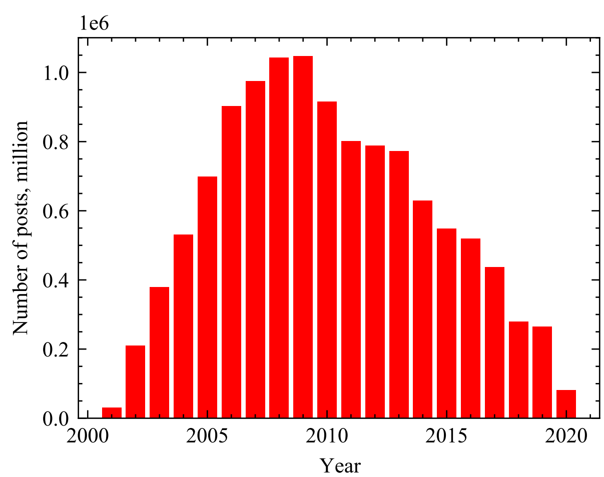


Figure 2. Distribution of posts by year.

The forum is the first dedicated to, and one of the most popular, right-wing online discussion forums [17]. The posts from the forum were collected using The Dark Crawler [18],

TABLE I
TOP TEN LANGUAGES IN THE FILTERED DATASET

Language	Number of posts
English	2,036,430
Italian	25,020
French	15,275
Spanish	14,192
Serbian	12,309
Portuguese	12,187
Croatian	5,091
German	3,187
Russian	2,111
Serbo-Croatian	943

and contain approximately 12 million posts spread over 1 million threads by almost 360,000 users. The full dataset consists of posts from the Stormfront forum from the 28th of August of 2001 to the 29th of April 2020. The dataset captured contains all thread and post information, containing post-date, author of the post, post content, thread and sub-forum information. Figure 2 illustrates the distribution of posts by year.

In this study, only data from 2015 onwards were considered. All posts containing only links, emojis, and those with less than five words were excluded since they do not tend to represent any topic but instead provide sentiments. Furthermore, the dataset contained posts created in several languages. Various languages in a corpus can confuse almost all the vocabulary-based NLP models because each language significantly enlarges a corpus's vocabulary. Therefore for each post from the resulting dataset, the language was identified using the fasttext model [19] [20]. Table I shows the top ten languages and corresponding numbers of posts written in each language. Only posts in English were used, resulting in a final total post count of 2,036,430.

V. RESULTS

A. Determining the number of topics

One of the hyperparameters for Latent Dirichlet Allocation is the number of topics. To find the best value, the perplexity metric was used. Perplexity for a topic modelling model shows an ability to generalise the results, that is, it shows how a model performs if it were to be applied to data, which was not used for training. For the experiment, the data were split into training and testing data sets, 75% and 25% respectively. Then LDA was trained on the training data, and applied to the test data with the following numbers of topics 5, 10, 30, 40, 50, 70, 100 and 200, we calculated perplexity for each of these values on the test set. The lowest perplexity value was for 40 topics, and thus was the value that was used in the following analysis.

B. Topics between January 2015 and April 2020

Once the model is trained and the hyperparameter value was selected, an exploratory analysis was performed to get preliminary insights from the matrix of the probability distribution of topics for each document. Each value in this matrix is the probability that the document belongs to the topic. Figure 3 demonstrates the average probability for each topic. The top three values of the average probability are for topics 30, 33 and 38, meaning that more posts tended to belong to these topics. Consequently, these topics are discussed more often than others in the forum. Table II shows the top fifteen keywords for these topics.

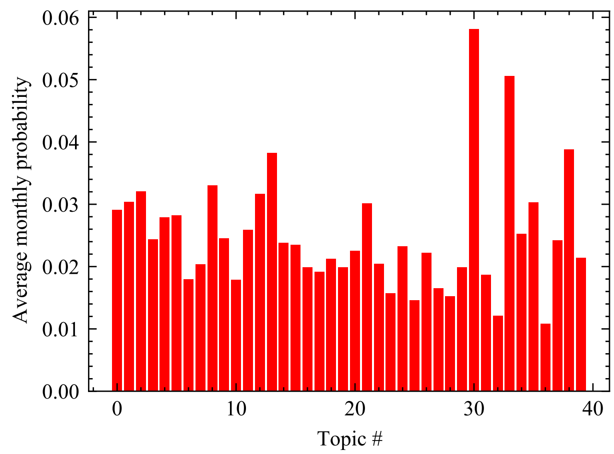


Figure 3. Average monthly topic probabilities.

TABLE II
TOP FIFTEEN KEYWORDS FOR TOPICS 30, 33 AND 38

Topic	Keywords
30	guy thing didn yeah lol thought negro crap guess hell white gay sick big funny
33	forum stormfront site find posting link comments didn google doesn internet article made lot comment
38	political party movement nationalist left support hope years real wn public members media politics change

The general picture about topic distribution showed how topics are distributed against each other, we wanted to identify individual topics and determine if they can be used to detect events, which happened from 2015 to 2020. Thus, monthly average probabilities were calculated for each topic. It was found that, as a way of presenting the level of discussion of a particular topic, if the average probability of a topic is higher in one period than in another then that topic was discussed more during that period. In other words, posts had a higher chance of belonging to this topic than to others; consequently it means that the topic appeared more frequently in this period. Figure 4 shows the comparison monthly average probabilities of the most discussed topic (number 30) and the least discussed (number 36). As shown on the plots, one line is higher than the other in all the points, meaning that topic 30 was always discussed more than topic 36 since 2015. Words on the top of each plot are the top keywords for the topic.

For example, a randomly selected post from topic 30 stated:

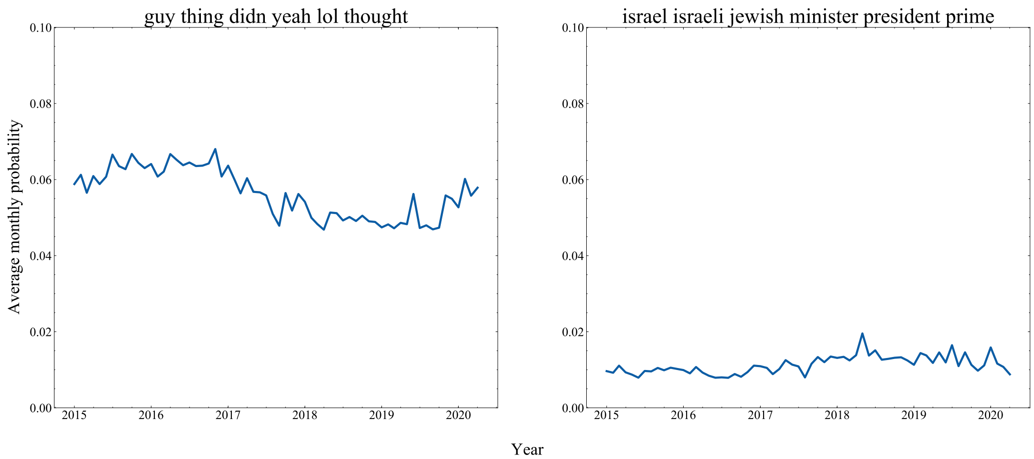


Figure 4. Average monthly topic probabilities for topics 30 and 36.

I was on my way back from the gym today when I could see two negro boys walking towards me. I knew, but absolutely knew, they were going to say something to me. As soon as they get within five feet of me they start grunting and looking at my shirt and reading the letters on it out loud. What is it with Negro kids? Whites never seem to do that kind of thing. It's like the Negro can't let anyone walk in peace. They HAVE to be noticed or affirmed. Even if it means embarrassing a complete stranger.

While a randomly selected post from topic 36 stated:

Quote: Bolivia plans to ask for a public meeting of the United Nations Security Council after U.S. President Donald Trump announces on Wednesday that the United States recognizes occupied Al-Quds (Jerusalem) as the capital of the Zionist entity and will move its embassy there. Bolivian U.N. Ambassador Sacha Sergio Llorentty Soliz said it would be a "reckless and a dangerous decision that goes against international law, the resolutions of the Security Council, it also weakens any effort for peace in the region and also upsets the whole region." Bolivia to Seek UN Security Council Meeting on Al-Quds (Jerusalem) Status – Al-Manar TV Lebanon

The probabilities of some of the topics have spikes, which indicate that it is likely that some disruption occurred around that time, reflecting the events in the real world. For example, Figures 5 and 6 show spikes in the second half of 2016, and it is likely that both topics are related to the 2016 United States presidential election [21], meaning that in the second half of 2016 the election-related these topics were being discussed a lot. This is a completely expected result, which validates the methods presented here.

Topic 16 contained a series of health-related keywords (ex: drug, medical, disease). Throughout 2015-2020, this topic was relatively consistent, and a relatively minor topic. However, at the beginning of 2020, discussions around this topic increased five-fold (Figure 7), exactly coinciding with the emergence of the COVID-19 pandemic [22].

In August 2017, the Stormfront website was banned by

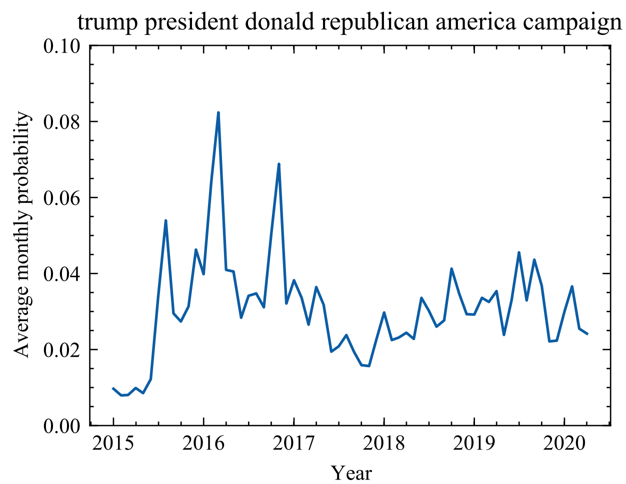


Figure 5. Average monthly topic probabilities for topic 2

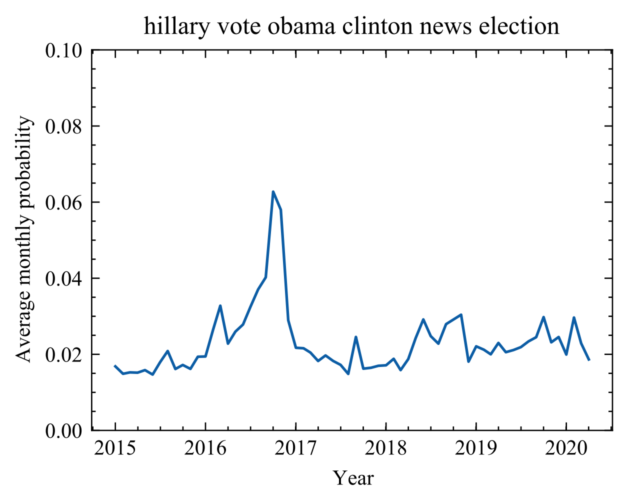


Figure 6. Average monthly topic probabilities for topic 24

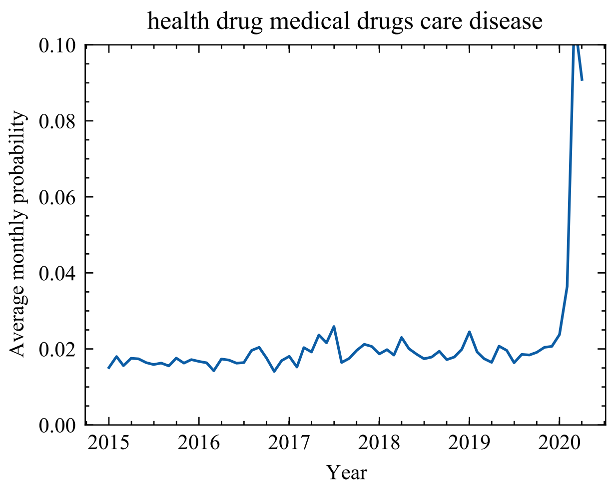


Figure 7. Average monthly topic probabilities for topic 16

its registrar. Figure 8 demonstrates the forum reaction to the event, where a very strong spike can be seen for the topic related to discussions around keywords “forum, Stormfront, site” in August 2017 [23].

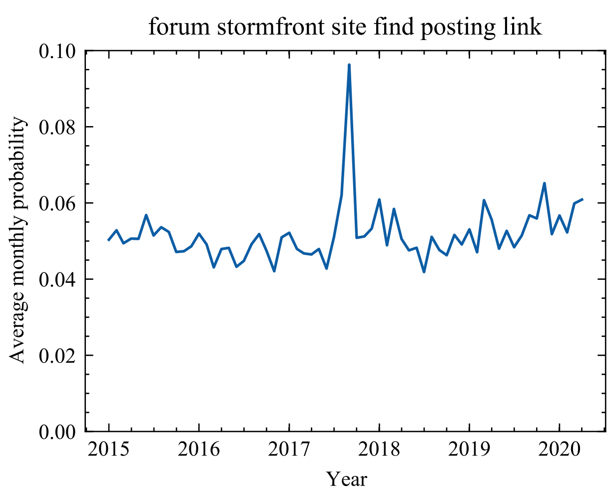


Figure 8. Average monthly topic probabilities for topic 33

Several other topics also had spikes in August 2017; therefore, they were analysed in more detail to check if there is a relationship among them. Figures 9 and 10 show the weekly data for 2017 for two topics; both of which increased in August 2017. The spike for the topic with keywords ”car shot dead fight” happens around week 33 (corresponding to the week of August 13, 2017) and most probably reflects discussions around Barcelona’s crowd rammings, which occurred August 17, 2017 [24]. Another topic with keywords ”hate speech media university” has posts related to the Unite the Right rally in Charlottesville, Virginia in August 2017 [25] and the Annual Stormfront Smoky Mountain Summit in September 2017. All the spikes in August-September 2017 have different natures, and thus it was concluded that the spikes are not related to each other.

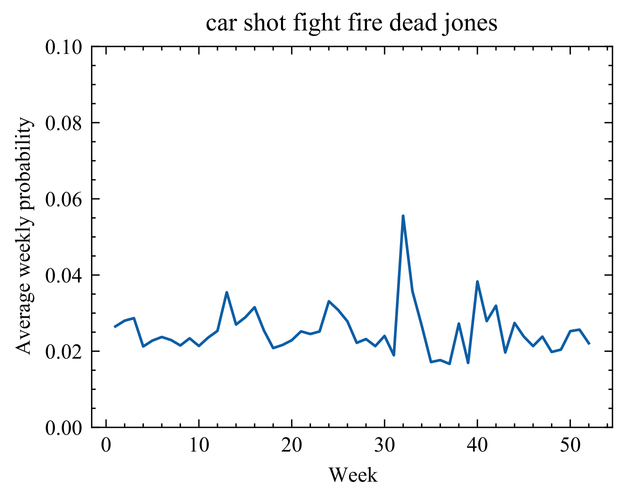


Figure 9. Average weekly topic probabilities for topic 3 in 2017

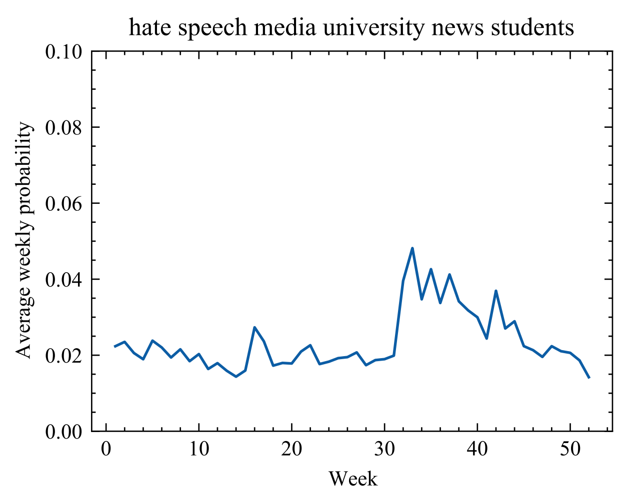


Figure 10. Average weekly topic probabilities for topic 19 in 2017

C. Weekly topics in 2019

The analysis of long periods is essential because it shows the general patterns, but the detection of smaller-scale events would allow security organizations to potentially detect quickly emerging threats or problems. Therefore data from January 2019 to May 2020 (non-inclusive) was selected, and the same trained model was applied. When all the topics were returned, the weekly average probabilities for each topic were calculated.

Figure 11 demonstrates the topic’s results with keywords ”military syria isis iran security” the peak in January 2020 coincides with a United States drone strike near Baghdad International Airport, resulting in an assassination of the major general Qasem Soleimani [26].

Figure 12 shows the topic with top keywords ”trump president donald republican america campaign”, the sudden

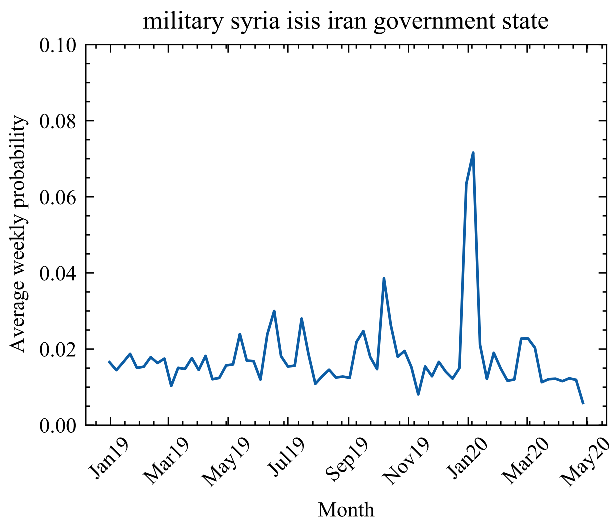


Figure 11. Average weekly topic probabilities for topic 23 from Jan 2019 to Apr 2020

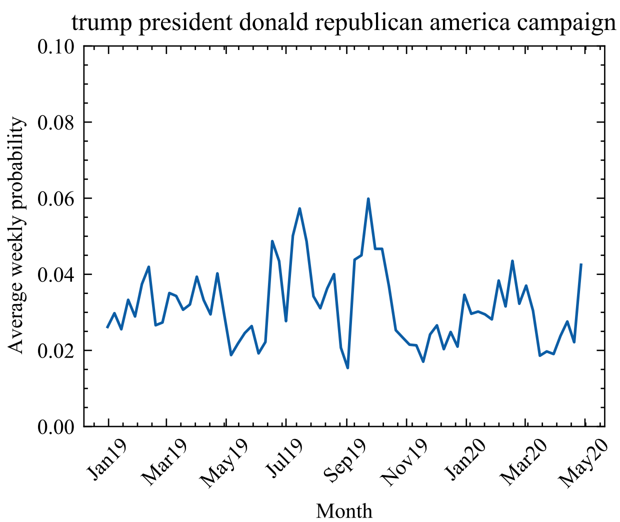


Figure 12. Average weekly topic probabilities for topic 2 from Jan 2019 to Apr 2020

increase in June 2019 is likely related with Trump's launch of his 2020 re-election campaign in Orlando, Florida [27]. Moreover, two peaks, at the beginning of July 2019 and end of September, coincide with his July 4th speech and with the 74th Session of the United Nations General Assembly, respectively.

VI. Conclusion

The models' and further analysis' results demonstrate that there is a correlation between real events and topics discussed on the forum; however, correlation does not imply causation, and thus we do not claim that the topics within the discussions lead to actions. We can see what specific topics have peaks at the time when something happens.

The distribution of topics (the vector of average probabilities of each topic within a certain period) can represent a snapshot of the community agenda, which in turn shows a general picture of the community members' attention. The model and further analysis of its results can show how the community reacts to the events - if it is a brief blimp, quickly reverting to average, then it disappeared from the agenda pretty quickly, see Figure 8, or if there is a sudden surge gradually decreasing to average after the event then it was an impactful event of great interest within the community, see Figure 5.

However, the proposed work has several areas of improvement. First of all, it is based on a bag-of-words representation of text, it works well for English because it is an analytical language, but if the intention is to use the model with more synthetic languages, for example Finnish, then changes are required, which help to reduce vocabulary of the corpus (for example, lemmatisation). Further, the model should be retrained from time to time to stay relevant, and finally, the model depends on some random processes so results can differ from run to run.

In conclusion, we can claim that anyone who analyses large-scale communities like StormFront - researchers, law enforcement, and security organisations - can use the proposed methods to detect changes in discussion topics, and detect new topics of interest.

References

[1] M. Conway, "Terrorist 'use' of the internet and fighting back," Information & Security: An International Journal, vol. 19, 01 2006, pp. 9–30.

[2] White homicide worldwide. Southern Poverty Law Center. 400 Washington Avenue, Montgomery, AL 36104. Apr. 1, 2014. [Online]. Available: https://www.splcenter.org/20140331/white-homicide-worldwide [retrieved: March, 2021]

[3] F. Bajak. Online providers knock 8chan offline after mass shooting. ABC News. Aug. 6, 2019. [Online]. Available: https://abcnews.go.com/Technology/wireStory/security-cut-off-cesspool-hate-8chan-forum-64778026 [retrieved: March, 2021]

[4] P. Gill, E. Corner, A. Thornton, and M. Conway, What are the roles of the Internet in terrorism? Measuring online behaviours of convicted UK terrorists. VOX Pol, 2015.

[5] P. Gill, E. Corner, M. Conway, A. Thornton, M. Bloom et al., "Terrorist use of the internet by the numbers: Quantifying behaviors, patterns, and processes," Criminology & Public Policy, vol. 16, no. 1, 2017, pp. 99–117.

[6] Homegrown violent extremist mobilization indicators 2019 edition. Federal Bureau of Investigation. Washington, DC: Office of the Director of National Intelligence. [Online]. Available: https://www.dni.gov/index.php/nctc-newsroom/nctc-resources/item/1945-homegrown-violent-extremist-mobilization-indicators-2019 [retrieved: March, 2021]

[7] H. Choi and H. Varian, "Predicting the present with google trends," Economic record, vol. 88, 2012, pp. 2–9.

[8] R. Lu and Q. Yang, "Trend analysis of news topics on twitter," International Journal of Machine Learning and Computing, vol. 2, no. 3, 2012, p. 327.

[9] B. Perry and R. Scrivens, "A climate for hate? an exploration of the right-wing extremist landscape in canada," Critical Criminology, vol. 26, no. 2, 2018, pp. 169–187.

[10] D. Dentice, "ṡo much for darwin⁼ an analysis of stormfront discussions on race." Journal of Hate Studies, vol. 15, no. 1, 2019, pp. 133–156.

[11] A.-M. Bliuc, J. Betts, M. Vergani, M. Iqbal, and K. Dunn, "Collective identity changes in far-right online communities: The role of offline intergroup conflict," New media & society, vol. 21, no. 8, 2019, pp. 1770–1786.

[12] R. Scrivens, "Exploring radical right-wing posting behaviors online," Deviant Behavior, 2020, pp. 1–15.

[13] R. Omizo, "Machine learning approaches to understanding white supremacy online," Rhet Ops: Rhetoric and Information Warfare (Eds. Jim Ridolfo and William Hart-Davidson), 2019, pp. 142–157.
[14] T. K. Landauer and S. T. Dumais, "A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge." Psychological review, vol. 104, no. 2, 1997, p. 211.
[15] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," the Journal of machine Learning research, vol. 3, 2003, pp. 993–1022.
[16] T. Hofmann, "Probabilistic latent semantic indexing," in Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, 1999, pp. 50–57.
[17] G. Bolaffi, R. Bracalenti, P. Braham, and S. Gindro, Dictionary of Race, Ethnicity & Culture. SAGE Publications, 2003, p. 254.
[18] R. Scrivens, T. Gaudette, G. Davies, and R. Frank, "Searching for extremist content online using the dark crawler and sentiment analysis," in Methods of Criminology and Criminal Justice Research. Emerald Publishing Limited, 2019.
[19] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," arXiv preprint arXiv:1607.01759, 2016.
[20] A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou et al., "Fasttext.zip: Compressing text classification models," arXiv preprint arXiv:1612.03651, 2016.
[21] D. C. Beckwith. United states presidential election of 2016. Encyclopedia Britannica. Nov. 11, 2019. [Online]. Available: https://www.britannica.com/topic/United-States-presidential-election-of-2016 [retrieved: March, 2021]
[22] D. B. Taylor. A timeline of the coronavirus pandemic. The New York Times Company. Jan. 10, 2021. [Online]. Available: https://www.nytimes.com/article/coronavirus-timeline.html [retrieved: March, 2021]
[23] J. Biggs. Another neo-nazi site, stormfront, is shut down. Verizon Media, Techcrunch. Aug. 28, 2017. [Online]. Available: https://techcrunch.com/2017/08/28/another-neo-nazi-site-stormfront-is-shut-down/ [retrieved: March, 2021]
[24] A. Smith, A. Jamieson, K. Simmons, and K. Itasaka. Spain terror: American among 14 killed in van and car attacks. NBC News. Aug. 18, 2017. [Online]. Available: https://www.nbcnews.com/news/world/van-hits-pedestrians-barcelona-n793506 [retrieved: March, 2021]
[25] H. Spencer and S. G. Stolberg. White nationalists march on university of virginia. The New York Times Company. Aug. 11, 2017. [Online]. Available: https://www.nytimes.com/2017/08/11/us/white-nationalists-rally-charlottesville-virginia.html [retrieved: March, 2021]
[26] M. Crowley, F. Hassan, and E. Schmitt. U.s. strike in iraq kills qassim suleimani, commander of iranian forces. The New York Times Company. Jan. 2, 2020. [Online]. Available: https://www.nytimes.com/2020/01/02/world/middleeast/qassem-soleimani-iraq-iran-attack.html [retrieved: March, 2021]
[27] M. Haberman, A. Karni, and M. D. Shear. Trump, at rally in florida, kicks off his 2020 re-election bid. The New York Times Company. Jun. 18, 2019. [Online]. Available: https://www.nytimes.com/2019/06/18/us/politics/donald-trump-rally-orlando.html [retrieved: March, 2021]

A Multiagent Meta-model for the Description of Socio-cognitive Processes:

An Enactive Perspective of Language in Artificial Agents

Luciana de Freitas *, Davidson Bruno da Silva *, Elayne de Moura Braga *,
Leonardo Lana de Carvalho * and Joao E Kogler Jr †
* *Federal University of the Valleys of Jequitinhonha and Mucuri, Diamantina, Brazil,*
Email: leonardolana.carvalho@ufvjm.edu.br
† *Polytechnic School of Engineering, University of Sao Paulo, Sao Paulo, Brazil, Email: kogler@lsi.usp.br*

***Abstract*—We propose here a meta-model based on the enactive and autopoietic theory for the description of social and language processes by Multiagent Systems (MAS). Our meta-model is a four-quadrant map, which describes, according to the concepts of structural couplings and organizational and functional closure, the architecture of the agents, as well as their social practices and the formation of social organizations. We propose a social cycle in which the architecture of agents evolves according to their practices, uses and social groups, the basis on which language evolves. Our meta-model was developed from two sources: a) the enactive and autopoietic theory of cognition, with an emphasis on social phenomena and the linguistic domain; b) a four-quadrant map, which describes MAS from an integral view. The motivation for the proposal of a new meta-model for the description of MAS and the emerging language in agent-based systems was the incompatibility between Wilber's integral view and that of the enactive theory about psychosocial processes. The central point of the enactive criticism that we carried out here was the concepts of interiority and individuality present in Wilber's theory.**

***Index Terms*—Cognition; Socio-cognitive processes; Enaction; Meta-model; Language; Multi-agent model**

I. Introduction

Meta-models are of great help in building models that involve very complex frameworks, typical in the study of socio-cognitive relations among agents and the environment, serving as scaffolding for the development of simulation architectures. The purpose of this article is to introduce a meta-model for the description of Multiagent Systems (MAS) and the emergence of language in artificial agents.

The proposed meta-model is derived from two other meta-models: a) that of the enactive and autopoietic theory of cognition, which describes language as a social phenomenon and the linguistic domain as coordination of behavior and consensual coordination of behavior [12]; [13]; [14]; b) the four-quadrant map based on the theory of Wilber [18], where each quadrant of the meta-model indicates a perspective in which artificial agents, their practices, collective actions and social organizations can be described and discussed [1]; [6]; [7]; [15]. The meta-model described by Maturana [12] and the meta-model describing an integral view on MAS, proposed by Ferber [6] both emphasize the description of language and other social processes as an emerging social practice in complex social organizations. However, the differences between the enactive and autopoietic view compared to Wilber's integral view [18] cannot be made compatible, as we will argue. The central point of our conceptual analysis rests here on the concepts of interiority and individuality present in Ferber's meta-model [6] that demanded an intense conceptual review from the enactive perspective, which resulted in the proposal of our meta-model for describing socio-cognitive processes, from a strictly enactive perspective. The concepts of interiority and exteriority have been reduced and replaced by the concepts of organizational and functional closure and; the concepts of individual and collective were analyzed, as is traditionally done in the enactive and autopoietic perspective, from the concept of structural coupling orders – levels – (first, second and third order structural couplings).

The paper is organized as follows: Section II presents the enactive and autopoietic theory of social and language phenomena. Following, section III presents the quadrangle map that describes the four aspects of a MAS, derived from a decomposition: Individual / Collective and Interior / Exterior. Section IV then presents our proposal for a meta-model to describe cognitive processes, such as language as a social phenomenon, using a four-quadrant map.

II. Enactive and autopoietic theory of social and language phenomena

Enaction is bring forth a world, an experiential world. In Enactive Cognitive Science, we argue that the realization of the experience occurs within the organizational and functional closure of structurally coupled living machines [12]; [13]. The figure 1 illustrates the history of recurring interactions of living beings, A and B, in the medium C. Maturana considers these recurring interactions as recursions, and not as simple repetitions in a sequence [12]. He gives an example of such recursive function: $\sqrt{a} = a'; \sqrt{a'} = a''; \sqrt{a''} = a'''$.

Then we say that A and B are organizationally and functionally closed. In biological systems, the organization of A is represented in the Maturana meta-model above and refers to the process of self-organization of the physiological domain. It is also said that A is functionally closed to medium C,

 ISBN: 978-1-61208-847-1

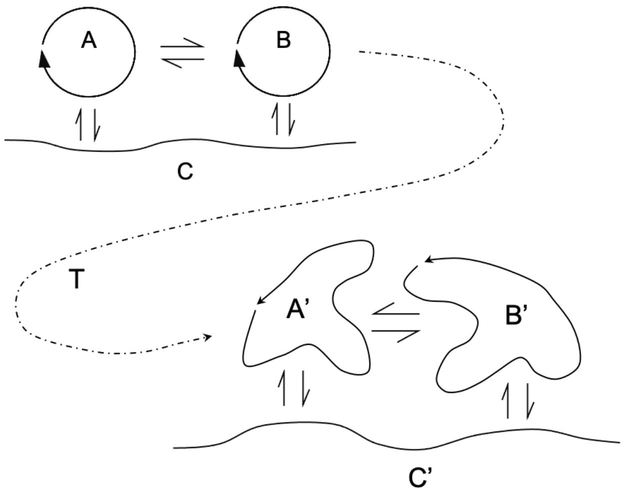


Fig. 1. History of interactions in a medium. A, B and C named as different systems by an observer change depending on their interactions while remaining congruent with themselves in the forms A', B' e C'. (Adapted from [12], p. 82)

altering the medium by its conducts and suffering medium disturbances. This is the domain of the behavior of system A. In the position of A, systems B and C are part of the medium. For B, the medium consists of A and C. Systems that self-produce structurally coupled to the medium, such as A and B, are said ontogenic phenotypes [12]. Notice that A, B and C change along the course of interactions, undergoing a set of transitions, collectively denoted T. The functional closures of A with B, A with C and B with C, result under T, respectively in the functional closures A' with B', A' with C' and B' with C'. According to Maturana [12], the observer notes this organizational change of A in A', B in B' and C in C'. Recurring interactions that occur in the contingencies of structural changes in the environment, undergo also structural changes due to their coupling to their respective local media. While dealing with living beings, the observer reckons a history of transformations (T) in the conducts, which constitute recurrent interactions. In a history of transformations, it is possible that there will be an expansion of dominance in the contingencies of changes. Consensual coordinations of conducts inaugurate the linguistic domain [12]; [13]. An example of language coordination is the situation that occurs when a couple starts their living with a first baby, and some friend visiting the house notices, as an external observer: "Wow, the dynamics of this house have completely changed!" The parents and the baby initiate a history of recurring interactions where the coordination of conduct between family members and the baby undergoes changes in their environment as a result of that history. A domain of consensual behavior is then established, as it was easily noticed by an observer.

Maturana emphasizes that language appears in this type of "co-drift story", "as an inevitable condition of the history of recurring interactions". The language emerges as "coordination of behavior and consensual coordination of behavior" ([12], p. 85), as a set of normative social practices [2]; [9]. An example of behavior in the linguistic domain is the friend's speech when visiting the newborn. It is classically highlighted in enactive and autopoietic theory that the emergence of language does not demand anything special, such as internal mental content or intentional representation, these concepts are reduced and eliminated by the concepts of functional closure [2]; [9]. Chalmers, also, highlights this point in his functional theory of language [4].

In the following section, we will go on to describe an integral view of MAS using a quadrants-based meta-model. In the section IV we will present the meta-model from an enactive perspective for describing MAS and the realization of social processes, like the emergence of language in MultiAgent Based Simulation (MABS).

III. Multiagent systems in an integral view: language as normative social practices

In order to describe different aspects of Multiagent systems, Ferber [6] leaned on Wilber's theory [18], who describes in four quadrants an integral view for psychology based on the concepts of subjectivity, objectivity, interobjectivity and intersubjectivity. The subjectivity quadrant is obtained by the Individual / Interior composition (I-I). The objectivity quadrant is obtained by the Individual / Exterior composition (I-E). The interobjectivity quadrant is obtained by the Collective / Exterior composition (C-E) and the intersubjectivity quadrant is obtained by the Collective / Interior composition (C-I).

Individual/Interior (I-I)	**Individual/Exterior (I-E)**
I *Subjectivity* <mental states, emotions, beliefs, desires, intentions, cognition> *"Interiority"*	**It, This** *Objectivity* <agent behavior, object, process, agent body> *"Observables, exteriority"*
Collective/Interior (C-I)	**Collective/Exterior (C-E)**
We *Intersubjectivity* <shared / collective knowledge, ontologies, norms and conventions, language and semantics> *"Noosphere"*	**Them, All This** *Interobjectivity* <reified social facts and structures, Organizations, forms of interaction, environment as interaction space> *"Sociosphere"*

Fig. 2. Analysis of MAS according to an integral view ([6], p. 15).

A series of psychological and social aspects are described in the meta-model, serving as a basis for understanding the investigated phenomena for the design of agent and MAS architectures. In Wilber's integral view, mental states, emotions, beliefs, desires, intentions and even cognition are described as subjective phenomena or as inherent to the interiority of individuals [18]. In this quadrant everything that concerns to the "I" (the "self") would be comprised. Behaviors of agents, of objects, observable processes and the bodies of agents are described as objectives or external things. In an emblematic way, this quadrant is described as being in the order of "It" or "This". In the integral view, facts and social structures are subject to reification. Organizations, forms of interaction and the environment as a space for interaction are

 ISBN: 978-1-61208-847-1

described as interobjective phenomena or as of the order of the “Sociosphere”. In this quadrant everything that is of the order of ”Them” or ”All This” would be comprised. Shared knowledge and collective knowledge, ontologies, social norms and conventions, language and semantics are described as intersubjective or of the order of the “Noosphere”, which is all about ”We” or ”us”.

In our view, quadrants are an interesting tool for the schematic understanding of popular descriptions of psychological and social phenomena. Mapping these descriptions is useful to transpose notions expressed about a psychosocial theme in models and simulations based on agents. Nevertheless, Wilber’s view is not compatible with the enactive and autopoietic theory of psychosocial phenomena [18]. In both meta-models, from Maturana [12] and Ferber [6], we find a description of language as social norms and practices. However, fundamental differences do occur, which we will be explored in the next section.

IV. A MULTIAGENT META-MODEL FOR THE DESCRIPTION OF SOCIO-COGNITIVE PROCESSES

The central point of our conceptual analysis rests on the concepts of interiority and individuality present in Ferber’s meta-model [6] that demanded a conceptual review from an enactive perspective, which resulted in the proposal of our meta-model below for describing MAS. From an enactive perspective, the concepts of interiority and exteriority are reduced and eliminated by the concepts of organizational and functional closure and; the concepts of individual and collective were analyzed as is traditionally done in the enactive perspective from the concepts of structural coupling of second and third orders.

Traditionally, agents are described as autonomous units of information processing, notion that derive from a cognitivist perspective [3]. According to Ferber, however, the agent is a computational entity, a process, located in a virtual or real environment [6]. Starting from the theoretical framework of the enactive and autopoietic theory, we defend that an artificial agent is defined first of all by its organization and then by its functioning in the environment. Thus, the agent is an organizationally and functionally closed unit, and artificial agents have, as designed by the human being, their initial organization and thus, an initial mode of functioning in the environment.

In the proposed meta-model it is intended to situate very common concepts in the domain of MAS. The agent and its architecture, which may or may not have a semiotic structure [10]; [11]; [16]; [17], is described as an organizationally closed unit, which when concretely implemented (as it is commonly said), will be structurally coupled by a human designer in a physical (material) level of the natural organization. In this sense, robots like Sophia from Hanson Robotics, are designed with such a body structure that enables it to be inserted in a certain way in the social dynamics of human relationships.

Starting with the upper-left quadrant of figure 3, from the composition “Structural coupling level n / Organizational closure” (SCn-OC), we describe the concepts of agent, agent architecture and semiotic structure of agent architecture. Following to the upper-right quadrant, from the composition “Structural coupling level n / Functional closure” (SCn-FC), we describe the behaviors of agents in the environment, the concepts of situated agent and embedded agent. These first two quadrants are central to the so-called Agent Centered Multiagent Systems (ACMAS) approach.

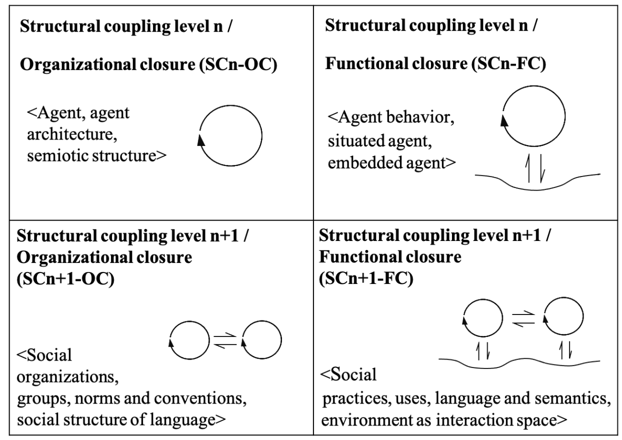


Fig. 3. Analysis of MAS according to an enactive vision. The figure illustrates language in artificial agents occurring from a semiotic structure, situated behaviors, social practices and a social structure of language, that is; as “coordination of behaviors and consensual coordinations of behaviors”. Social practices occur as behavior coordinations, but consensual coordinations depend on the formation of social organizations; social norms and conventions and, then, the emergence of a social structure of language.

The following two quadrants at the bottom of figure 3, describe fundamental characteristics in the so-called Organization Centered Multiagent Systems (OCMAS) approach. In this case, the concept of emergent organization takes the focus of the design and implementation of the MAS [6]. From the composition “Structural coupling n+1 / Functional closure” (SCn+1-FC), there is a series of emerging phenomena. Classically, it is understood that micro-behaviors (at level n) can lead to macro-behaviors (at level n+1). We then describe in the bottom left quadrant social practices, uses, the semantics as uses, the language as coordination of the practices of agents in their environments and the environment as a space modifiable by interactions. In the last quadrant, at bottom right, we zoom in again on the internal dynamics of the system, highlighting the formation of social groups, with a social organization being a group of at least one social group [6]. At this level of the organization of the system we describe social organizations, social norms and conventions and also the structure of language. It is emphasized that these phenomena cannot be reduced to the processes described in the architectures of the agents. The social processes in their organizations provide normative patterns of disturbances on the structure of the agent, which may lead to the evolution of their architecture. The following meta-model in the figure 4 presents the evolution of these ontologies described above.

With the modification of the agent’s architecture due to the

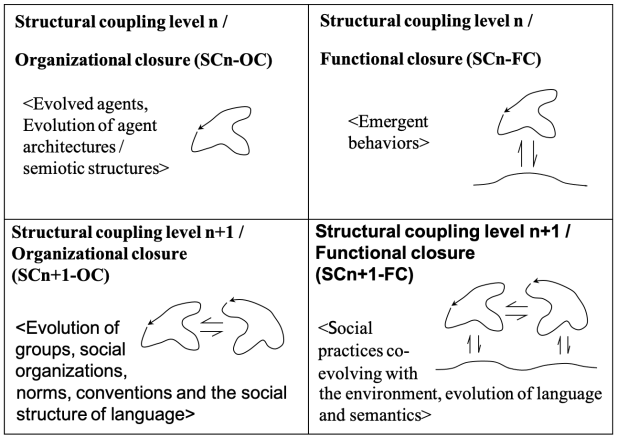


Fig. 4. Evolution of MAS according to an enactive vision.

emergence of normative social practices embodied in social organizations, a new cycle begins in the MAS. Thus, the new architecture will provide for the emergence of behaviors in the system, social practices will evolve together with the environment, leading to the evolution of language and semantics (emergent uses in MAS). Finally, the social organization itself also evolves as a result of these collective practices, with the possibility of forming new social groups, norms and conventions, and thus also of the social structure of language. We understand that language variation occurs fundamentally with the formation of different social groups, it occurs as variations in the norms of social practices.

We understand to capture here the essentials of Maturana and Varela views about language as "history of co-drift", "as an inevitable condition of the history of recurring interactions". The language emerges as "coordination of behavior and consensual coordination of behavior" [12]; [13]; [14], as a set of normative social practices. The realization of language in artificial agents is the result of several recurring functions in MAS defining the agent architecture, as well as the MAS architecture. The organizational and functional closure embodied in certain structural coupling level leads to coordinations of behaviors (collective practices), and consensual coordinations of behaviors (normative social practices), occurring from an agent architecture (that can have a semiotic structure), situated behaviors, social practices and a social structure of language. Social practices appear from coordinations of behavior, however, consensual coordinations require the formation of social organizations, with social norms and conventions and, then, with the realization of the social structure of language. However, a challenge is imposed in the design of semiotic architectures of agents from an enactive perspective. As a requirement, it is necessary that the semiotic structure of the agents can be self-developed by the system, not directly implemented by the programmer.

V. Final considerations

The objective of this article was to describe social processes and among them language in particular, based on: a) the enactive and autopoietic theory of cognition, which describes language as a social phenomenon and the linguistic domain as processes of communication over communications [12]; [13]; [14]; b) the Ferber's meta-model [6], which describes MAS from a view based on Wilber's [18], where each quadrant of the meta-model describes psychosocial processes from the subjectivity and objectivity dichotomy. Fundamental aspects of MAS, such as the architecture of agents, the formation of artificial organizations and the emergence of language were presented from this meta-model [1]; [6]; [7]; [15].

However, even though the enactive theory and the integral view describe language as a social process, occurring in social organizations, based on social structures and processes, there are incompatible differences between the enactive view and the Wilber's integral view. Fundamental ontological differences were pointed out, which led us to propose a meta-model to describe MAS and the realization of language in artificial agents in a strictly enactive perspective, using a four-quadrant map. The central point of our conceptual analysis highlighted that the concepts of interiority and individuality present in Ferber's meta-model [6] are liable to be reduced and replaced by the concepts of organizational and functional closure and that, the concepts of individual and collective were analyzed as it is traditionally done in the enactive perspective from the concepts of structural coupling of second and third orders.

We conclude with the understanding that our meta-model of description of MAS under an enactive vision captures the essentials of Maturana and Varela about language as a history of co-drift [12]; [13]; [14], "as an inevitable condition in the history of recurring interactions" (p. 85). We emphasize that language emerges as "coordination of behavior and consensual coordinations of behavior" ([12], p. 85), as a set of social practices in different forms of social organizations. However, a challenge is imposed in the design of semiotic architectures of agents from an enactive perspective because it is necessary that the semiotic structure of the agents architecture can be self-developed by the Multiagent system and not directly implemented by the designer, which is a target for a future work.

References

[1] L. L. Carvalho, "Emergencia de Linguagem a partir do ciclo social da regra em modelos computacionais," Anais da 43a Reuniao Anual da Sociedade Brasileira de Psicologia Brazil, ISSN 2176-5243, October 2013 [43a Reuniao Anual da Sociedade Brasileira de Psicologia Brazil, October 2013].

[2] L. L. Carvalho and J. E. Kogler, "The enactive computational basis of cognition and the explanatory cognitive basis for computing," Cognitive Systems Research, vol. 67, pp. 96–103, June 2021.

[3] L. L. Carvalho, D. J. Pereira and S. A. Coelho, "Origins and evolution of enactive cognitive science: Toward a enactive cognitive architecture," Biologically Inspired Cognitive Architectures, v. 16, pp. 169–178, April 2016.

[4] D. A. Chalmers, "Computational Foundation for the Study of Cognition" Journal of Cognitive Science, vol. 12(4), pp. 323–357, 2011.

[5] D. A. Chalmers, "The Varieties of Computations: A Reply" Journal of Cognitive Science, vol. 13(3), pp. 211—224, 2012.

[6] J. Ferber, "Multi-agent concepts and methodologies," in Agent-based modelling and simulation in the social and human sciences, D. Phan and F. Amblard, Eds. Oxford: Bardwell Press, 2007, pp. 7–33.

[7] J. Ferber, L. L. Carvalho, D. Phan and F. Varenne, ”O que e seguir uma regra? Reflexao sobre as normas e os usos,” in Temas em Ciencias Cognitivas e Representaçao Mental, E. J. Lopes, Ed. Porto Alegre: Sinopsys, 2012, pp. 259–277.
[8] J. Ferber, F. Michel, J. BAEZ, ”AGRE: Integrating Environments with Organizations,” E4MAS’04: Environments for Multiagent Systems Australia, pp. 127–134, 2005.
[9] D. D. Hutto, E. Myin, A. Peeters and F. Zahnoun, ”The Cognitive Basis of Computation: Putting Computation in Its Place,” in The Routledge Handbook of the Computational, M. Sprevak and M. Colombo, Eds. London: Routledge, 2018, pp. 272—282.
[10] A. C. Loula, R. Gudwin and J. Queiroz, ”Synthetic approach of symbolic creatures,” S.E.E.D. Journal – Semiotics, Evolution, Energy, and Development, vol. 3(3), pp. 125–133, 2003.
[11] A. C. Loula, R. Gudwin and J. Queiroz, ”Symbolic communication in artificial creatures: an experiment in Artificial Life,” Lecture Notes in Computer Science, vol. 3171, pp. 336-345, 2004.
[12] H. Maturana, ”Cogniçao, ciencia e vida cotidiana,” 2nd ed., Belo Horizonte: Editora UFMG, 2014.
[13] H. Maturana and F. J. Varela, ”Autopoiesis and Cognition: the Realization of the Living,” Dodrecht: Reidel, 1980.
[14] H. Maturana and F. J. Varela, ”A Arvore do Conhecimento: As bases biologicas do entendimento humano,” Campinas: Psy, 1995.
[15] D. Phan, J. Ferber, L. L. Carvalho, F. Varenne, ”Qu’est-ce que suivre une regle ? Reflexion sur les normes et les usages,” in Ontologies pour la modelisation par Systemes Multi-Agents en Sciences Humaines et Sociales, D. Phan, Ed. Paris: Hermes-Sciences et Lavoisier, 2014.
[16] L. Steels, ”Language as a Complex Adaptive System,” Parallel Problem Solving from Nature PPSN VI, pp. 17–26, 2000 [International Conference on Parallel Problem Solving from Nature PPSN, 2000].
[17] L. Steels, ”Intelligence with representation,” Phil. Trans. Roy. Soc. London, vol. A361(1811), pp. 2381–2395, 2003.
[18] K. Wilber, ”Theory of Everything,” Boston: Shambala, 2000.