COGNTIVE 2025

The Seventeenth International Conference on Advanced Cognitive Technologies and Applications

ISBN: 978-1-68558-260-9

April 6 - 10, 2025

Valencia, Spain

COGNITIVE 2025 Editors

Muneo Kitajima, Nagaoka University of Technology, Japan

COGNITIVE 2025

Forward

The Seventeenth International Conference on Advanced Cognitive Technologies and Applications (COGNITIVE 2025), held on April 6 – 10, 2025, 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
- Autonomous systems and autonomy-oriented computing
- Hot topics on cognitive science

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 2025 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 2025. 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 2025 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope COGNITIVE 2025 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. We also hope that Valencia provided a pleasant environment during the conference and everyone saved some time to enjoy this beautiful city.

COGNITIVE 2025 General Chair

Jaime Lloret Mauri, Universitat Politecnica de Valencia, Spain

COGNITIVE 2025 Steering Committee

Jayfus Tucker Doswell, The Juxtopia Group, Inc., USA
Thomas Ågotnes, University of Bergen, Norway
Muneo Kitajima, Nagaoka University of Technology (Emeritus), Japan

COGNITIVE 2025 Publicity Chair

Francisco Javier Díaz Blasco, Universitat Politècnica de València, Spain
Ali Ahmad, Universitat Politècnica de València, Spain
José Miguel Jiménez, Universitat Politècnica de València, Spain
Sandra Viciano Tudela, Universitat Politècnica de València, Spain

COGNITIVE 2025

Committee

COGNITIVE 2025 General Chair

Jaime Lloret Mauri, Universitat Politecnica de Valencia, Spain

COGNITIVE 2025 Steering Committee

Jayfus Tucker Doswell, The Juxtopia Group, Inc., USA
Thomas Ågotnes, University of Bergen, Norway
Muneo Kitajima, Nagaoka University of Technology (Emeritus), Japan

COGNITIVE 2025 Publicity Chair

Francisco Javier Díaz Blasco, Universitat Politècnica de València, Spain
Ali Ahmad, Universitat Politècnica de València, Spain
José Miguel Jiménez, Universitat Politècnica de València, Spain
Sandra Viciano Tudela, Universitat Politècnica de València, Spain

COGNITIVE 2025 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
Thangarajah Akilan, Lakehead University, Canada
Piotr Artiemjew, University of Warmia and Masuria in Olsztyn, Poland
Divya B, SSNCE, India
Thierry Bellet, University Gustave Eiffel, France
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
Jerome Dinet, University of Lorraine, France
Serena Doria, University G.d'Annunzio Chieti, Italy

Table of Contents

Metacognition-Driven Preprocessing for Optimized Artificial Intelligence Performance

Naavya Shetty
Undergraduate in Computer Science and Philosophy
Department of Philosophy
University of Illinois Urbana-Champaign
Illinois, United States
e-mail: leltuz404@gmail.com

***Abstract*—Machine cognition is currently heavily speed-based. Directly tackling inputs with computation often leads to inefficient steps, such as performing redundant or repetitive computation, or execution without assessing whether a task is within computational capacity. This paper proposes a preprocessing metacognitive system to be implemented in a manner such that it screens all input requests, creating a strategic 'bottleneck' to filter, redirect or halt the flow of control before computation begins. The findings theorise improved accuracy, reliability and resource-management, strengthening the argument for making metacognition an essential component of artificial intelligence.**

Keywords-artificial intelligence; resource optimization; self-modifying machines.

I. Introduction

Formulating as accurate a response in as little time as possible is the goal computation strives to achieve, but there is no system in place to determine whether it has the computational ability to do so, nor to overcome redundant operations, thus failing to optimize processing efficiency. Implementing solutions to these issues requires a kind of 'awareness' in computation, which poses challenges in terms of defining self-assessment standards, developing algorithms to monitor computational efficiency, and integrating self-adapting decision making processes. Machines lack the concept of cognitive overload, making it difficult to ascertain when an operation should be adjusted or entirely terminated before execution. There also exists the issue of balancing computational overhead with the benefits of self-regulation.

Existing research in this space such as Schaeffer [1] primarily focuses on detecting suboptimal actions in various forms of machine learning contexts. The following research, however, aims to extend this by developing a preprocessing metacognitive system that not only identifies inefficiencies but also proactively consults a database of prior experiences to optimize computational resources. This broader scope addresses not just action evaluation but also strategic planning and resource management, offering a more comprehensive solution to the efficiency of artificial intelligence systems.

The need for a solution to this problem lies in the lack of computational efficiency leading to wasted resources, increased latency, and suboptimal decision-making, which could lead to cascading inefficiencies or altogether failure, especially in high-stakes applications such as autonomous systems and large-scale simulations. Addressing this gap by introducing a metacognition-based system could enable the existing infrastructure to allocate resources more dynamically, recognize when to rethink strategies, and improve performance while simultaneously minimizing computation.

The core scientific problem is the absence of metacognition in current artificial intelligence software, preventing systems from evaluating their computational strategies and optimizing efficiency. Unlike human cognition, which involves self-regulation and resource allocation, current systems lack the mechanism to assess when an operation is redundant or inefficient. The challenge lies in developing an architecture that enables such a system to recognize its own performance constraints and adjust computational processes without excessive overhead.

Through this paper, the definition of metacognition will be analyzed and mapped into machine cognition in such a way that it sheds light on and provides impetus to a possible adjacent feature of artificial general intelligence that can enable evaluation of its own computational limits, streamline decision-making, and reduce inefficiencies in problem-solving. What is sought here is a form of automative 'wisdom' over intelligence. In other words, the goal is to augment and improve decision-making abilities in machines, in addition to being able to deliver intelligent responses.

Some questions that will be tackled through this paper are how metacognition can be modeled and implemented in artifical intelligence systems to enhance preprocessing strategies that can enable self-evaluation of computational efficiency, what mechanisms allow machines to adjust preprocessing strategies dynamically, how machines can detect inefficiencies or redundant operations in data preprocessing, and how metacognitive preprocessing improves the adaptability and robustness of current artificial intelligence systems.

The purpose of this paper is to explore how metacognitive principles can be integrated into artificial systems to enhance their decision-making efficiency. By equipping such a system with a form of self-awareness regarding its computational processes, the research aims to reduce redundant operations, optimize resource allocation, and improve overall system intelligence. This work lays the foundation for models that not only generate intelligent responses but also assess and refine their reasoning processes.

This proposal is not without its limitations. Implementing metacognition in computation introduces computational overhead, which could paradoxically reduce efficiency if not carefully managed. Additionally, defining an objective metric for 'effort' in machine cognition remains an open challenge.

While biological systems can optimize through evolutionary processes, artificial systems lack intrinsic motivation, making it difficult to determine when computational adjustments are necessary. Furthermore, existing architectures may require fundamental modifications to accommodate self-regulation mechanisms effectively.

This paper first delves into the background of artificial intelligence systems and metacognition, as well as some relevant definitions for the proposed theory. It then describes the proposed theory, describing the features of such a theory, the requirements for building it, and highlights necessary qualities in the implementation. Finally, the paper discusses some rebuttals that may be raised and provides an answer to them.

II. Background

Current endeavors towards building artificial intelligence are geared towards creating systems that can emulate and surpass human intelligence. The Dartmouth meeting of 1956, where the roots of artificial intelligence can be traced, established certain goals: [2]

> The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.

When such goals could be achieved, the emergence of independent machine intelligence was assumed to be the logical next step. Most research in the sphere of artificial intelligence has an anthropomorphized approach; this is fair to say considering that some of the more popular work is being done in neural networks, the foundation and organizational structure of which emulates how neurons interact in the human brain. While the human brain is admittedly not optimized for intelligence, meaning the forms of intelligence exhibited by it are not necessarily the best and most efficient forms vetted evolutionarily, it is also true that there are several aspects of human intelligence that would do well to be implemented in intelligent agents. One such feature is metacognition.

Plato spoke of a system that allows for a learner, a teacher and an evaluator within our cognitive system, making this the first possible mention of a third aspect that evaluates our cognitive processes beyond simply learning and thinking. This was later defined as 'metacognition' or the "knowledge and cognition about cognitive phenomena" [3]. In current literature, this is generalized to 'thinking about thinking', allowing "one's own beliefs, reasonings, desires, intentions… [as well as] cognitive abilities, motivational dispositions, practical reasoning strategies" [4] to determine the extent and accuracy of one's thoughts and decisions.

Some definitions need to be briefly touched upon before establishing the theory. The 'dual process theory' (DPT) accounts for the processing of cognition through two different processes, referred to as System 1 and System 2. System 1 is the immediate response to a situation that "operates automatically and quickly, with little or no effort and sense of voluntary control" [5]. On the other hand, System 2 is the more deliberate system of thinking that "allocates attention to the effortful mental activities that demand it, including complex computations" [5]. The three significant distinctions between the two systems lies in:

1) the time between input and output,
2) the number of neurons activated in human cognition (here, we do not give in to chauvinism and instead allow for the generalization of 'neurons' to the smallest unit of cognitive ability, whatever that may be, depending on the agent of cognition. In the context of philosophy of mind, chauvinism refers to certain mental states being linked with physical elements limited to agents in which the state is confirmed [6]), and
3) the amount of energy required i.e., the entropy of the processes.

These three distinctions become crucial in the forthcoming argument of the need and implementation of metacognition.

It is essential to note here that this paper does not deal with the debate between DPT and single-process theory, and that it is only the features that are of import to the theory. Single process theorists believe that the difference arises due to degrees of cognition and that they are not separate kinds of processes altogether. Even so, both sides acknowledge the difference in terms of the three factors established above. Since resolving the debate "will not inform our theory development about the critical processing system underlying human thinking" [7], implying that beyond observing the outcome, there is nothing further to be understood about the architecture of cognition from establishing one over the other, this paper will instead proceed with a functionalist perspective in its use of DPT terminology. In the context of philosophy of mind, the concept of functionalism follows that "what makes something a mental state of a particular type does not depend on its internal constitution, but rather on the way it functions, or the role it plays, in the system of which it is a part" [8]. In this case, by taking a functionalist approach to System 1, we focus solely on the properties of such a system – properties which are observations of both processing theories regardless of their mechanisms – and therefore can be referred to by either theory with the same effect.

III. Proposal

The main focus of the approach in this paper is to significantly improve computational efficiency through two key features set up and handled by the metacognitive system:

1) Accessing a Database of Previous Computations:
   - *Creates and utilizes a centralized repository of past computations and problem-solving processes* - When encountering a new prompt, the system can search this database to find similar problems or computational patterns. This allows for faster resolution by reusing solutions, reducing the need for extensive recalculations.

 ISBN: 978-1-68558-260-9

- *Identifies the boundaries of expertise* - This is a critical extension of the strategy. In cases where a new prompt falls outside the range of knowledge or available solutions, the system can flag it and either redirect it to a more suitable computational resource or escalate it to human intervention. This ensures that the system remains efficient by focusing on problems within its scope while also handling edge cases appropriately.
- *Enables adaptive learning* - New computations can be refined or improved upon based on the knowledge and data collected from previous tasks, further optimizing the system's performance over time. This approach can also involve the redirection of prompts to relevant components or their respective field-specialized units, ensuring effective use of computational resources.

2) Segmenting Computation:
   - *Focuses on utilizing prior computations that share similar structures, properties, or nature* - By recognizing patterns and similarities between past and current tasks, computations can be reused instead of being repeated from scratch.
   - *Allows for parallel processing* - By breaking down tasks into specialized components, multiple processes can run concurrently, improving speed and efficiency without overwhelming any single, or the overall, available computational resource.

At the very outset, we must consider the kind of machine this theory can be implemented best through. Metacognition requires strong analysis tools that identify patterns, something best modelled by deep learning models. Schaeffer [1] integrates metacognitive processes into reinforcement learning frameworks. His model demonstrates that metacognition can be algorithmically implemented, enabling machines to detect their own suboptimal actions without external input. This provides a concrete example of how metacognitive functions can be translated into computational algorithms. Similarly, the proposed metacognitive system can be implemented via reinforcement learning, or other such machine learning methods, and can still be envisioned as independent of any computational cognitive system it is implemented in tandem with. The theory this paper proposes, the preprocessing metacognitive system (PMS), is the integration of the principles of metacognition and System 1 processes in computation, such that it may be able to provide a quick, immediate response before either proceeding in a certain manner or terminating computation altogether.

System 1, as the immediate response to an input, is heavily dependent on intuition in humans and "valid intuitions develop when experts have learned to recognize familiar elements in a new situation and to act in a manner that is appropriate to it" [5]. As such, intuition – and by extension, System 1 processes – can be reduced to an outcome of analysis, categorization and recognition of patterns which then manifests as a quicker response to situations that the untrained eye/mind would not notice patterns or decision-prompting cues in. More interestingly, however, is that this is only the case where one has expertise in the field, and that "when the question is difficult and a skilled solution is not available…[one] often answers an easier one instead, usually without noticing the substitution" [5]. This is a very human quality, a broad class of System 1 responses called 'heuristics' that are shortcuts established to answer questions with speed, regardless of accuracy. As mentioned before, speed-over-accuracy preference is also a persisting feature of machine cognition today.

If System 1-like processes were plainly implemented in machines, it would lead to false positives or false negatives, something that deep learning models are not exempt from. Like how people answer a particularly difficult situation almost immediately by looking at other cues that may not be relevant to the situation at hand, citing their intuition, machines also identify and call upon such misleading shortcuts when they are trained to observe patterns. This is most apparent when they mislabel or make errors in categorization based on certain other visible cues. Such models are widely implemented in deep learning models looking to identify images i.e., visual inputs and even in those, it is quite a challenge to train deep learning models to overcome these false results, requiring the dedication of many resources and an extensive database. To overcome these issues, the PMS becomes necessary. As previously mentioned, metacognition analyzes computational ability when it looks for patterns in reasoning to consider the matter of accuracy. Not only will it make it capable of terminating computation altogether if it falls outside the scope of such ability, but it will also carefully consider hindering features. Integration with System 1 principles of heuristics which are based on reasoning and skill-based cues allows it to create immediate output without resorting to actual computation, as well as avert the use of computation in cases that fall outside the ability of the machine altogether. This is what leads to trustworthy and robust outputs that only resolve things within the scope of the machine.

Thus, by acting as a system that makes use of analyzing tools before computation begins, the PMS can not only analyze the extent of, and patterns in computational ability to implement the benefits of a System 1 process – that of time, resource and energy efficiency – but also be able to determine the accuracy of the computation.

When looking into mapping the two systems onto each other, their inputs and outputs in the process become noteworthy. Metacognition acts as a separate system from cognition. The relation of DPT with metacognition can be understood such that the "default reasoning [System 1] is reasoning that precedes metacognitive control and intervening reasoning [System 2] is reasoning that follows metacognitive control" [9]. Let us consider the matter of what System 1 accepts as inputs and how its output correlates to the intermediary metacognitive system. System 1 accepts a task – be that identification, recognition, computation – objectively. On the other hand, metacognition processes 'reasoning cues' produced by System 1 while it develops judgements about the task.

At the outset, it is necessary to note that mapping one onto the other does not call for replacing one with the other but instead a strategic mixture of their features, as has been necessitated above. If the metacognitive process were to be mapped onto a System 1 process, there would be a restructuring of I/O, and processing. It would develop such that the I/O of the PMS would be of metacognitive nature i.e., reasoning cues give directions and judgements, while the process itself follows the System 1 architecture, making it a matter of objective analysis of the cues and making connections using established heuristics to develop the directions as outputs. To further establish this idea, we consider the fact that the main computation, which will be of System 2-like nature, with deliberate control and use of more emphasized resources, can accept directional input from a boolean perspective, and have its own input passed on to it. Thus, to generate robust outcomes, the system would require a restructuring of the procedural hierarchy and not a replacement. That said, metacognitive systems would require their reasoning cues to come from somewhere, and this would come directly from an increasing database of the computation itself. The cycle would proceed as such: the metacognitive system is an 'onlooker' observing how computation works in a deliberate System 2 fashion. In the initial stages, it will be completely passive, only building a foundation of what kinds of problems the computation system faces, what kind of inputs generate a certain kind of input and what this can say about the processing ability of the machine. These will become the cues it calls upon when faced with new inputs, gradually identifying patterns in them and dealing with it accordingly.

Currently, most machine intelligence research targets actual computation, by accepting the input as is and working on it. This is a reasonable way to proceed when developing intelligence, but the implementation of this integrated system would allow for enhanced and accurate outputs by filtering out most inputs that have either already been computed before or fall outside the scope of ability, thus only utilizing the System 2-like deliberate computation to deal with new, solvable problems. It is also important to note that this system would not possess any cognitive abilities of its own but is entirely a pattern-recognition system. It identifies the requirements of the input and either matches it to a certain heuristic that has been established over repeated computations or gives a negative output altogether. In both cases, the system requires access to observe the computation itself of the machine it is working with – the data it works with comes from the machine itself and not from any external source of data. This is a similar process to how humans develop metacognition to form the basis of judgements, beliefs, reasoning, etc.

IV. Discussion | Evalution

Certain rebuttals may be raised against this theory; some of these will hereby be addressed. Perhaps the foremost one is the matter of metacognition not being a necessary component of human cognition, questioning its necessity in machine cognition. While it may be true that humans do not always rely upon metacognitive systems, the fact remains that it is full of erroneous judgements that develop from an oft flawed System 1 response. This arises due to the lack of logical integrity in the formation of heuristics. The principles of metacognition make it possible for the additional component of accuracy that is derived specifically from patterns analyzing computational ability.

The necessity of metacognition in current computation may still be doubted because there exist models today that function smoothly even without it, but it must be considered that none of these models have truly managed complex intelligence, and that some of the best work remains language models and neural networks. To reach a stage of cognition several orders of magnitude higher than the present, at the level of artificial general intelligence, there are certain other peripheral features that become necessary to ensure efficiency and accuracy, one of which is metacognition. Its necessity is based solely on the fact that intelligent machines require careful handling of resources, a 'wisdom' that develops through long-term analysis. The idea is like that of hierarchical, version-based intelligence, and could perhaps do away with the need of several versions of AGI if it can become a separate, constantly learning system entirely responsible for efficiency while the computational power increases separately. In simpler terms, wisdom deepens while intelligence improves.

There exist some common issues with models that are based on repeated learning, such as overfitting, where they provide excellent results with training data but fail with test data, because of overly adhering to certain features in the training sets and incorporating elements that would otherwise be deemed noise. While systems like regularization do exist to handle such issues, the prime issue with implementing such a system in PMS would be that it sacrifices accuracy for generalization ability. Instead, emphasis can be given to the idea of false positives and false negatives that were discussed previously. To expound upon the idea, the directions that will be given as output from the PMS will direct attention to the features of the input that hinder accurate computation, either due to lack of history in such spheres, or due to lack of clarity. In either case, the way it is handled by the computation once again becomes a learning for the PMS, allowing it to deal with such inputs differently and more effectively further on.

V. Conclusion and Future Work

To summarize, this paper explores how artificial intelligence can enhance its efficiency through the implementation of a preprocessing metacognitive system that acts as an automatic first-line-of-defense for all computation requests. It discusses how the proposed system can dynamically adapt its preprocessing strategies based on context, improving efficiency and accuracy in data handling. The system assesses its own preprocessing steps, identifies inefficiencies, and adjusts accordingly, mimicking human-like reflective thinking. This approach aims to enhance the adaptability, robustness, and self-improvement of further artificial intelligence systems in much more complex environments. The main aspects this paper furthers in terms of existing literature are:

- Self-Evaluation Mechanisms: Implementing algorithms that allow for assessing the quality of computations and decisions

- Redundancy Detection: Developing systems that reference a database of previous computations to identify and eliminate redundant operations
- Adaptive Processing Strategies: Creating frameworks that enable AI to adjust its computational strategies based on real-time self-assessment, optimizing resource allocation

This paper allows for there to be further objections to the conceptual explanation and details of the PMS; however, it also seeks to establish firmly the need for this theory. Without such a system, while progress will be made, it is less likely to be as immediately efficient and fast developing as it could be than if it were with it.

Further work includes developing and testing algorithms with the aforementioned features, drawing from the reinforcement learning models tested by Schaeffer [1], as well as expanding the range and examining other deep learning and neural network related systems in order to test for the most efficient implementation of the PMS. By integrating reinforcement learning insights and exploring diverse neural network architectures, future work aims to enhance the PMS's performance, ensuring its robustness and efficiency across a wider range of applications.

REFERENCES

[1] R. Schaeffer, "An algorithmic theory of metacognition in minds and machines", *ArXiv https://arxiv.org/abs/2111.03745*, 2021.
[2] J. McCarthy, M. L. Minsky, N. Rochester, and S. C. E., "A proposal for the dartmouth summer research project on artificial intelligence", *https://raysolomonoff.com/dartmouth/boxa/dart564props.pdf*, 1995.
[3] J. H. Flavell, "Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry", *American Psychologist*, 1979.
[4] J. Greco, "The social value of reflection", *In W. J. Silva-Filho & L. Tateo (Eds.), Thinking About Oneself: The Place and Value of Reflection in Philosophy and Psychology. Springer International Publishing*, 2019.
[5] D. Kahneman, *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2011.
[6] B. Gertler, "In defense of mind-body dualism", *In R. Shafer-Landau & J. Feinberg (Eds.), Reason and Responsibility.*, 2007.
[7] W. De Neys, "On dual- and single-process models of thinking", *Perspectives on Psychological Science*, 2021.
[8] J. Levin, "Functionalism", *In E. N. Zalta & U. Nodelman (Eds.), The Stanford Encyclopedia of Philosophy (Summer 2023). Metaphysics Research Lab, Stanford University*, 2023.
[9] A. R. Dewey, "Metacognitive control in single- vs. dual-process theory", *Thinking and Reasoning*, 2023.

 ISBN: 978-1-68558-260-9

Implementation of Structured Memes into Behavioral Ecology via GOMS

Muneo Kitajima
Nagaoka University of Technology
Nagaoka, Niigata, Japan
Email: mkitajima@kjs.nagaokaut.ac.jp

Makoto Toyota
T-Method
Chiba, Japan
Email: pubmtoyota@mac.com

Jérôme Dinet
Université de Lorraine
Nancy, France
Email: jerome.dinet@univ-lorraine.fr

Katsuko T. Nakahira
Nagaoka University of Technology
Nagaoka, Niigata, Japan
Email: katsuko@vos.nagaokaut.ac.jp

***Abstract*— Our daily actions are executed to achieve desired states. Perceiving our own situation, we select actions that are expected to bring about the desired states, and execute them as a series. The memory used in this process is a representation in the brain of memes that are inherited from generation to generation. Memes are structured into three levels: action level, behavioral level, and cultural level, and are acquired through mimetic behavior. Memes at the higher levels are acquired as one gets older. This study is based on the Model Human Processor with Real-Time Constraints (MHP/RT), a cognitive architecture that includes Perceptual, Cognitive, and Motor (PCM) processes, and a memory system that is used during action selection by the PCM process and updated after action execution. We examine how the cognitive process of Two Minds utilizes memes structured in three layers, which is referred to as C-resonance in MHP/RT. It is known that knowledge built as a result of iterative actions toward a goal state is represented by a GOMS hierarchical structure whose elements are goals (G), operators (O), methods (M), and selection rules (S). This study shows that GOMS bundles memes belonging to different levels and combines goals and selection rules at the conscious level with methods and operators at the unconscious level to achieve effective and efficient goal-oriented action execution. The expressed behavior can be regarded as the result of crossing the syntax expressed by GOMS with the semantics expressed by memes, showing distinct characteristics depending on the balance of dominance between unconscious and conscious behavior in the behavioral ecology.**

Keywords*— *GOMS; Behavioral ecology; Meme; Resonance.

I. INTRODUCTION

In interacting with the environment for living, humans develop by selecting and executing actions and accumulating execution results while operating a cyclic loop of the Perceptual, Cognitive, and Motor (PCM) processes. The basis of action selection is imitation. The things – any cultural entity including objects and events that an observer might consider a replicator of a certain idea or set of ideas – to be imitated exist as memes; they are repeatedly imitated from generation to generation.

A. *Meme Proposed by Dawkins [1]*

The mechanism by which cultures and civilizations produced by humans are passed on from generation to generation was not clear. From the standpoint of cultural anthropology, Dawkins organized his research on the mechanisms of cultural inheritance. As a result, he argued that cultural inheritance cannot be explained solely in terms of the capability of memory on the part of humans, and that there must be a hypothetical existence on the part of culture that might convey information, such as genes. He coined the term "meme" for this indefinite virtual entity [1]. This idea itself received a lot of support, but the time passed without the mechanism being clarified [2].

When Dawkins proposed the meme, the function of genes was not yet understood. Therefore, Dawkins' explanation and others had many problems inherent in them due to misunderstandings about genes. Certainly, genes were replicators. However, they did not play the role of duplicating the blueprint of the finished product as conventionally thought, but rather, they played the role of plotting the process of growth that established the basic functional structure and its relationships. It was this role of the genes that enabled humans to be highly adaptable.

B. *Redefining Memes through their Connection to the PCM Process*

The memes proposed by Dawkins can be redefined by considering them as mappings of the individual's memory (which can be called *the individual ecological memes*), which is activated in the process of selecting and executing actions, onto *the collective ecology* that caries the culture. Memes are realized in the memory of each individual. They hold the relationships between events, which enable humans in an ever-changing environment to express effective behavior in each situation in generic forms that are valid across generations [3]. More specifically, the spatial coordinates and absolute times that characterize events occurring in the real world are *not* retained in the memes; they are dynamically determined according to the state of the environment when the behavior is expressed according to the representation of the memes. Figure 1 shows a schematic representation of the action selection and execution process in the real environment as the PCM process. In the perceptual process, humans perceive the state of the real environment through parallel processing of

 ISBN: 978-1-68558-260-9

	P: Perceptual Process		***C: Cognitive Process (Memory)***		***M: Motor Process***	
	⇒ Hearing		⇒ Meme_1			
	⇒ Vision		⇒ Meme_2			
Environment	⇒ Smell	⇒ Integration	⇒ Meme_3	⇒ Integration	⇒ Behavior ⇒	**Environment**
	⇒ Touch	***P-Resonance***	$\cdots$	***C-Resonance***		
	⇒ Taste		⇒ Meme_{N-1}			
			⇒ Meme_N			

Figure 1. PCM process and memes

the five senses and integrate what they perceive individually by binding them. In the cognitive process, the memes related to the perceived information are activated in parallel, and they are integrated as a series of operators that can be executed as concrete actions in the environment. In the motor process, the operators are executed through feedforward processing, keeping pace and synchronizing with changes in the environment in an unconscious manner.

C. Problem Statement

The PCM process runs synchronously with the environment, whereas, the "memory system" used by the PCM process updates itself asynchronously with the environment to reflect the results of the PCM process. Supported by the PCM process and the memory system, each individual repeats action selections in an ever-changing environment without any breakdown. In this regard, it is important to clarify the interface between the PCM process, which operates synchronously with the environment, and the memory system, which is not required to synchronize with the environment but is connected to the PCM process. The mechanism for connecting what is perceived with memory has been described as P-resonance [4]. Within the memories, structured with memes as elements, activation propagates in parallel with the integrated perceptual information as the activation source. Connecting activated memories to the actions performed in the real world can be rephrased as "enabling the activated memes in the real world by integration." In Figure 1, this mechanism is shown as C-resonance. How is this done?

When we unravel the origin and evolution of life, we can find a clue to the solution. Life is formed under the structures shaped by the atmosphere, oceans, energy cycles, and gravity that characterize the Earth, a planet in our solar system, spinning on its own axis and orbiting the Sun. The direction of life's evolution is determined by the pressures exerted by these structures. Life is formed as an adaptive body with the functional and structural feature that work most efficiently in the environment. It is best captured by the four elements of Goals, Operators, Methods, and Selection rules (GOMS) [5]. In this study, we show that C-resonance, which integrates the memes activated in parallel as the effective actions in the real world, can be explained by the GOMS concept.

D. Organization of the Paper

This paper is organized as follows. Section II describes the PCM process and memes, referring to our previous work. Section III describes the GOMS theory presented by Card, Moran, and Newell in their book entitled "The Psychology of Human-Computer Interaction," and describe the mechanism of C-resonance. Section IV discusses the characteristics of the behavioral ecology that emerge from the interaction of the structured meme content and the C resonance by GOMS. Section V summarizes this study and discuss its implications for the digital generation.

II. PCM Process and Memes

In this section, we explain the details of the PCM process, memories, and memes outlined in Figure 1, based on our cognitive architecture "Model Human Processor with Realtime Constraints (MHP/RT)" [6][7].

A. PCM Process and Resonance for Linking with Memory

When interacting with the environment, humans respond to physical and chemical stimuli emitted from the external and internal environment by sensory nerves located at the interface with the environment and take in environmental information in the body. The brain acquires environmental information concerning the current activity of the self through the multiple sensory organs. Further, it generates bodily movements that are suitable for the current environment. The stable and sustainable relationship between the environment and the self is established through continuous coordination between the activity of the self and the resultant changes in the environment, which should affect the self's next action.

Figure 2, adapted from [4, Figure 1], shows the process, based on MHP/RT [6][7], by which environmental information is taken into the body via sensory nerves, processed in the brain, and then acted upon by the external world via motor nerves. This process involves memory, which is modeled as Multi-Dimensional Memory Frame, and perceptual, cognitive (Two Minds), and motor processes. The memory structure, Multi-Dimensional Memory Frame, consists of Perceptual-, Behavior-, Motor-, Relation-, and Word-Multi-Dimensional Memory Frame. Perceptual-Multi-Dimensional

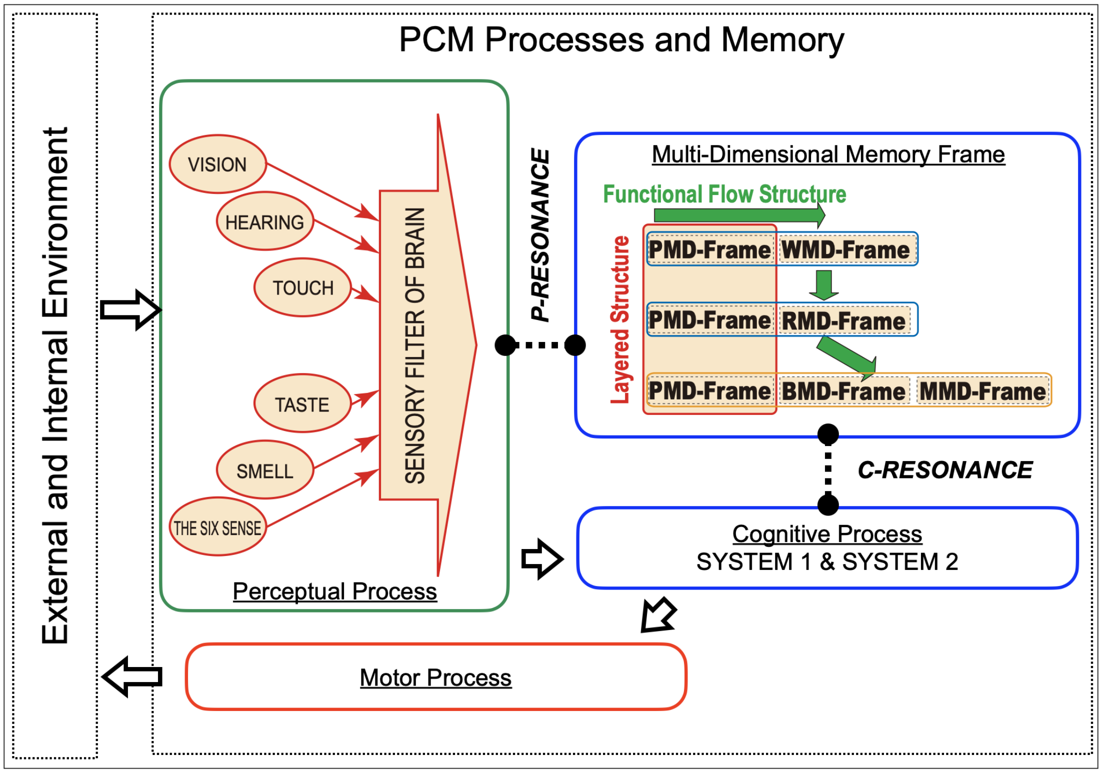


Figure 2. Information uptake by perceptual processes from the external and internal environment memory activation and execution of cognitive and motor processes through resonance [4, Figure 1].

Memory Frame overlaps with Behavior-, Relation-, and Word-Multi-Dimensional Memory Frame, for spreading activation from Perceptual- to Motor-Multi-Dimensional Memory Frame.

Perceptual information taken in from the environment through sensory organs resonates with information in the memory network structured as Multi-Dimensional Memory Frame, that is, P-Resonance. In Figure 2, this process is indicated by •—•. Resonance occurs first in the Perceptual-Multi-Dimensional Memory Frame and activates the memory network. After that, the activity propagates to the memory networks that overlap the Perceptual-Multi-Dimensional Memory Frame, which are Behavior-, Relation-, and Word-Multi-Dimensional Memory Frame, and finally to the Motor-Multi-Dimensional Memory Frame. In cognitive processing by Two Minds, conscious processing by System 2, which utilizes the Word- and Relation-Multi-Dimensional Memory Frame via C-Resonance, and unconscious processing by System 1, utilizing the Behavior- and Motor-Multi-Dimensional Memory Frame via C-Resonance, proceed in an interrelated manner. The motor sequences are expressed according to the Motor-Multi-Dimensional Memory Frame, which is the result of cognitive processing. The memories involved in the production of actions are updated to reflect the traces of its use process and influence the future action selection process.

B. P-Resonance Connecting Perceptual Processes and Multi-Dimensional Memory Frame

Information from the environment is taken into the brain via multiple sensory organs. The sensory organs are distributed throughout the body. In addition, the information received by sensory organs is time-series information. Therefore, the information received by sensory organs is spatially and temporally distributed. The brain integrates this disparate sensory information in some way, perceives it, and passes it on to cognitive processing. The question of how this integration is performed is known as the binding problem. We proposed that P-resonance provides a solution to the binding problem and showed the existence of basic senses that enable orderly processing of information from sensory organs. The basic senses include rhythmic sense related to time, spatial sense related to spatial perception, and number sense [4].

C. Memory as Structured Meme

When the PCM process is running, the contents of Perceptual-Multi-Dimensional Memory Frame are updated in response to the perceptual process, those of Word-, Relation-, and Behavior-Multi-Dimensional Memory Frame are updated in response to the cognitive process, and those of Motor-Multi-Dimensional Memory Frame are updated in response to the motor process. Figure 2 characterizes the memories of PCM process, i.e., the Multi-Dimensional Memory Frame, as the traces of its operation and classifies it into five sub-memories, i.e., Perceptual-, Word-, Relation-, Behavior-, and Motor-Multi-Dimensional Memory Frame. In short, it is an expression of the way in which the memories are structured, focusing on the continuous updating of memory associated

 ISBN: 978-1-68558-260-9

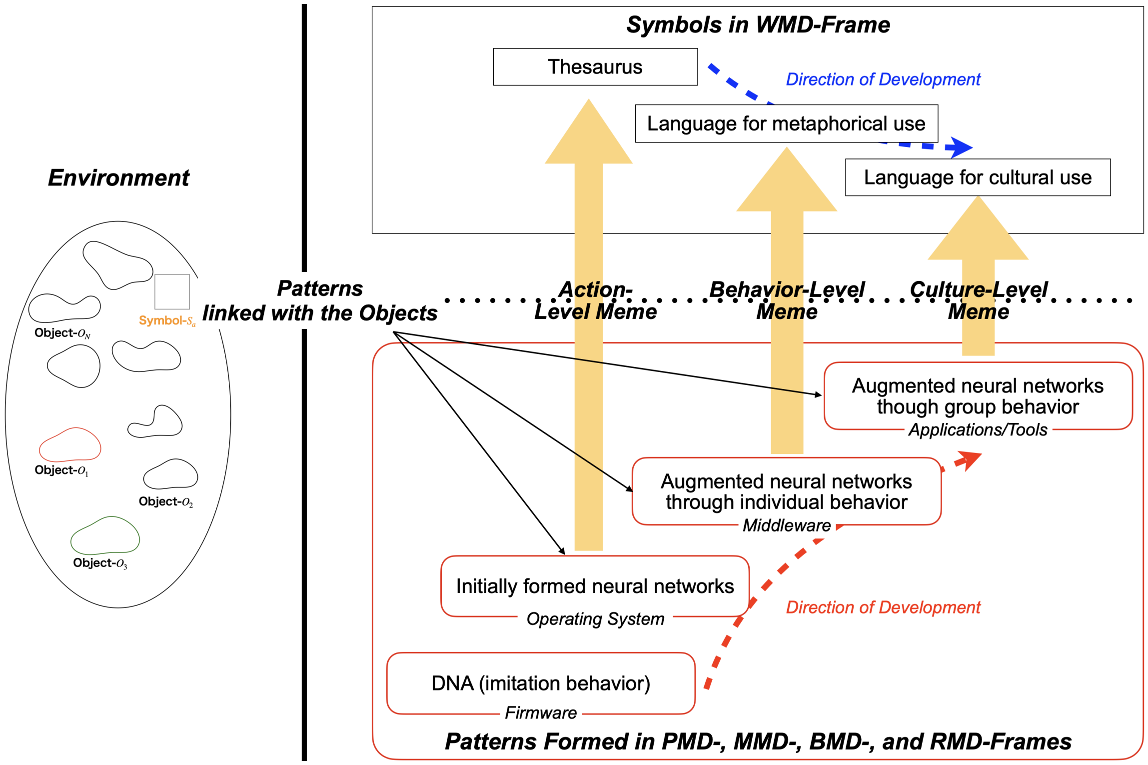


Figure 3. Structure of meme (adapted from [3])

with the execution of PCM process. It is important to note that, in the Multi-Dimensional Memory Frame, Perceptual-Multi-Dimensional Memory Frame overlaps with Behavior-, Relation-, and Word-Multi-Dimensional Memory Frame, for spreading activation from Perceptual-Multi-Dimensional Memory Frame to Motor-Multi-Dimensional Memory Frame.

Alternatively, the memory system can be viewed from the perspective of memory use. The integrated sensory information through the basic senses first activates the Perceptual-Multi-Dimensional Memory Frame (P-resonance); then the activation propagates to the Word-, Relation-, and Behavior-Multi-Dimensional Memory Frame, and finally to the Motor-Multi-Dimensional Memory Frame bound to the motor nerves. This process is repeated in an environment that changes from moment to moment, and satisfactory behavior is expressed in the environment. The basis of behavior is imitation, and what is imitable is limited according to growth stage. In addition, behaviors that can be imitated across generations are preserved as sustainable behaviors. In this way, we can organize the Multi-Dimensional Memory Frame, which is used by the PCM processes and updated by their execution, in terms of memes that can be inherited across generations [8].

Figure 3 shows a functional classification of the regions of the Multi-Dimensional Memory Frame that are activated by P-resonance after an object in the environment has been perceived. "Words" are considered the archetype of meme [9]. Words, i.e., symbols, in the Word-Multi-Dimensional Memory Frame are gradually incorporated into the environment in the form of *thesauruses*, i.e., lists of words in groups of synonyms and related concepts, languages used for person-to-person communication, *individual languages*, which might include not only direct but also metaphorical uses, and languages used in cultural contexts, *cultural languages*, 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 Word-Multi-Dimensional Memory Frame" 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, respectively, that were introduced in the Structured Meme Theory proposed by Toyota et al. [8]. Hereafter, the memes classified into Action-Level Meme, Behavior-Level Meme, and Culture-Level Meme are abbreviated by A-memes, B-memes, and C-memes, respectively.

The culture that exists in the environment is the integration of the C-memes that exist in the brain of an individual across all members of the group to which that individual belongs. This corresponds to the meme proposed by Dawkins. Blackmore, one of the theoretical followers of Dawkins' meme theory, argues that "memes are symbolized by the act of imitation" after examination of the theory of the meme [10]. This argument is consistent with the idea of memes as we have presented it, since we can think of it as saying that the A- and B-memes are physical and provide a stable basis for imitation, and that the C-memes above them is not mentioned because it is strongly dependent on the environment.

The mechanism by which the three levels of memes, and genes as well, inherit information is analogous to an information system. Genes serve as firmware that mimics behavior-level activities. A-memes serve as the operating system that defines general patterns of spatial-temporal behavioral functions. B-memes represent middle-ware that extends the general patterns to concrete patterns. C-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 the Multi-Dimensional Memory Frame are as follows:

- A-memes represent bodily actions stored in the Motor-Multi-Dimensional Memory Frame,
- B-memes represent behaviors in the environment stored in the Behavior-Multi-Dimensional Memory Frame, and
- C-memes represent culture stored in the Relation- and Word-Multi-Dimensional Memory Frame.

Meme-based behaviors, i.e., mimicry behaviors, are implemented in the real environment. Since what can be imitated depends on the individual growth stage, there are qualitatively different sets of mimicry behaviors to emerge depending on the growth stage. The bases of those mimicry behaviors are represented as A-memes, B-memes, and C-memes.

As Dawkins proposed, taking a meme-centric view of cultural inheritance is in itself essential. A person's genes express a memory resonance response mechanism. Through this mechanism, replications (resonance replications) are generated when there is a common experience. Memes are present in the environment as those that can cause these resonance replication. Such memes can be called cultures. Memes influence the phenotype called culture, but the resonance itself is formed as something unique in a person's experience. The fact that imitation is personal and influenced by environmental conditions does not guarantee that the phenotype will be perfectly imitated. This mechanism is common to the idea of ecological inheritance theory, named niche (ecological status) construction, advanced by Odling-Smee et al. in evolutionary ecology [11]. This mechanism makes it possible for humans to be highly adapted to their environment.

III. C-RESONANCE VIA GOMS

A. *Binding Problem at the Cognitive Level*

In Figure 3, the objects in the environment are shown to activate the A-, B-, and C-memes. In this activation process, various regions related to the objects are activated. In Figure 2, the propagation of activation within the Multi-Dimensional Memory Frame is shown as the functional flow structure. The expression is such that the activity propagates from the Perceptual-Multi-Dimensional Memory Frame to Word-, Relation-, Behavior-, and Motor-Multi-Dimensional Memory Frame in that order. However, the layers below the Word-Multi-Dimensional Memory Frame are not structurally overlapped. Therefore, the activity propagates layer by layer from the top to bottom via the Perceptual-Multi-Dimensional Memory Frame that overlaps with them; at the top, there is an activation flow from the Word- to Perceptual-Multi-Dimensional Memory Frame, at the middle from the Perceptual- to Relation-Multi-Dimensional Memory Frame, and at the bottom from the Perceptual- to Behavior-, finally to Motor-Multi-Dimensional Memory Frame. The portions of Word-, Relation-, and Behavior-Multi-Dimensional Memory Frame that are activated in this manner may contain multiple regions that may be related via the Perceptual-Multi-Dimensional Memory Frame but not directly related to each other.

Memories that hold memes are activated in parallel to be used by the PCM process, which is a serial process. Here, we can see another binding problem occurring at the cognitive level. In Figure 2, the bridge between cognitive and memory processes is shown as C-resonance for resolving the meme binding problem. The cognitive process might operate *carefully* by using the entire areas of the Multi-Dimensional Memory Frame that are activated in connection with the Perceptual-Multi-Dimensional Memory Frame. The advantage of this method is that reality can be guaranteed by referring to the contents active in the Perceptual-Multi-Dimensional Memory Frame. However, it is *inefficient* because it is an interpreter-like process. At the perceptual level, the binding problem of perceptual information is solved by P-resonance for effective use of the perceptual information. At the cognitive level, C-resonance resolves the problem of efficient use of memory by binding memes somehow that are activated in parallel [12]. What is the equivalent of the basic senses in P-resonance in C-resonance?

B. *GOMS*

1) GOMS as a Meta-Structure for Understanding Behavioral Ecology: Humans select actions that contribute to the realization of the state they desire to achieve. The principles at work in the execution of the cognitive activity of action selection are the bounded rationality and satisficing principles [13]. Such action selection processes are modeled by the serial firing of procedural knowledge expressed in the form of production rules, "IF conditions are satisfied, THEN perform actions [14][15][16]." Individual action choices are expressed in terms of firing sequences of production rules, which are procedural knowledge. However, if we take a bird's-eye view of the situations that each individual encounters, the firing sequences of procedural knowledge applied in more or less similar situations will show certain patterns. Card, Moran and

 ISBN: 978-1-68558-260-9

Activation of Memes via P-Resonance			Utilization of Memes via C-Resonance	
Perceptual Process	Multi-Dimensional Memory Frame	Memes	GOMS	Cognitive Process
	Word- & Relation-	Culture-Level	Goals and Selection Rules	System 2
Basic Senses	Behavior-	Behavior-Level	Methods	System 1
	Motor-	Action-Level	Operators	System 1

Figure 4. Relation between GOMS, memes, and Two Minds

Newell [5] identified GOMS as concepts that represent such patterns.

GOMS specifies concepts that define a meta-structure that is essential for understanding the ecology of human behavior. Aristotle's theory of the four causes was the first theory to systematize such a meta-structure. Allen Newell et al. reconstructed it from a cognitive scientific perspective and constructed the GOMS theory.

2) Definition of GOMS: GOMS is an analytic technique for making quantitative and qualitative predictions about skilled behavior with a computer system. GOMS is defined as follows (adapted from [5, Chapter 5, pp.144–146]):

> The user's cognitive structure consists of four components: (1) a set of Goals, (2) a set of Operators, (3) a set of Methods for achieving the goals, and (4) a set of Selection rules for choosing among competing methods for goals. We call a model specified by these components a GOMS model.
> ***Goals***. A goal is a symbolic structure that defines a state of affairs to be achieved and determines a set of possible methods by which it may be accomplished.
> ***Operators***. Operators are elementary perceptual, motor, or cognitive acts, whose execution is necessary to change any aspect of the user's mental state or to affect the task environment.
> ***Methods***. A method describes a procedure for accomplishing a goal. It is one of the ways in which a user stores his knowledge of a task. The description of a method is cast in a GOMS model as a conditional sequence of goals and operators, with conditional tests on the contents of the user's immediate memory and on the state of the task environment.
> ***Selection Rules***. When a goal is attempted, there may be more than one method available to the user to accomplish the goal. The selection of which method is to be used need not be an extended decision process, for it may be that task environment features dictate that only one method is appropriate. On the other hand, a genuine decision may be required. The essence of skilled behavior is that these selections are not problematical, that they proceed smoothly and quickly, without the eruption of puzzlement and search that characterizes problem-solving behavior. In a GOMS model, method selection is handled by a set of selection rules. Each selection rule is of the form "if such-and-such is true in the current task situation, then use method M."

Behavioral goals are represented by a robust hierarchical structure. There is a primary behavioral goal, G, and underneath it, there are subgoals, G′, that must be accomplished to complete the primary goal, and then there are sub-subgoals, G′′, to complete the individual subgoals. The final node that unfolded the task is the operator. One node above it is the method, and one level above it is the node representing the selection rule. The goal structure from top to bottom looks like "G-G′-G′′ · · · S-M-O."

3) Binding Memes via GOMS: In GOMS, behavioral goals form a robust hierarchical structure, and the goal structure mediates the organization of behavior. Achieving a goal, G, requires achieving the subgoals underneath it, G′s. This structure does not hold the time as its primary parameter. The order between G′s is important. The time elapsed for executing G′ is associated with the operators located at the bottom layer, which connect to the motor process of PCM that implements the contents of Motor-Multi-Dimensional Memory Frame, i.e., the operators, in the real world.

On the other hand, the mechanism of action execution based on MHP/RT is explained as follows. As shown in Figure 2, the environment is perceived and connected to the Multi-Dimensional Memory Frame by P-resonance. Then, as shown in Figure 1, the memes having been acquired by structuring the Multi-Dimensional Memory Frame through experience are activated, and the A-memes are connected to the real world to execute the action. As mentioned earlier, in the functional flow structure within the Multi-Dimensional Memory Frame shown in Figure 2, behavior generation following the flow of activity through the Perceptual-Multi-Dimensional Memory Frame is inefficient. It is reasonable to assume that GOMS is used to structure A-, B-, and C-memes that do not contain absolute temporal and spatial information as a method of realizing behavior generation without breaking down, while keeping in sync with the real world where the situation changes from moment to moment. GOMS should correspond to the phenomenon of A-, B-, and C-memes binding without the Perceptual-Multi-Dimensional Memory Frame when encountering certain situations, indicating the entity of the phenomenon of C-resonance. This may correspond to a shortcut that may be formed within the Multi-Dimensional Memory Frame.

Figure 4 shows the correspondence between memes and GOMS. Among the activated memes, the combinations of C-, B-, and A-memes that have formed GOMS bonds in the process of gaining experience are processed by System 2 and System 1, and the operator sequence is executed in the real

 ISBN: 978-1-68558-260-9

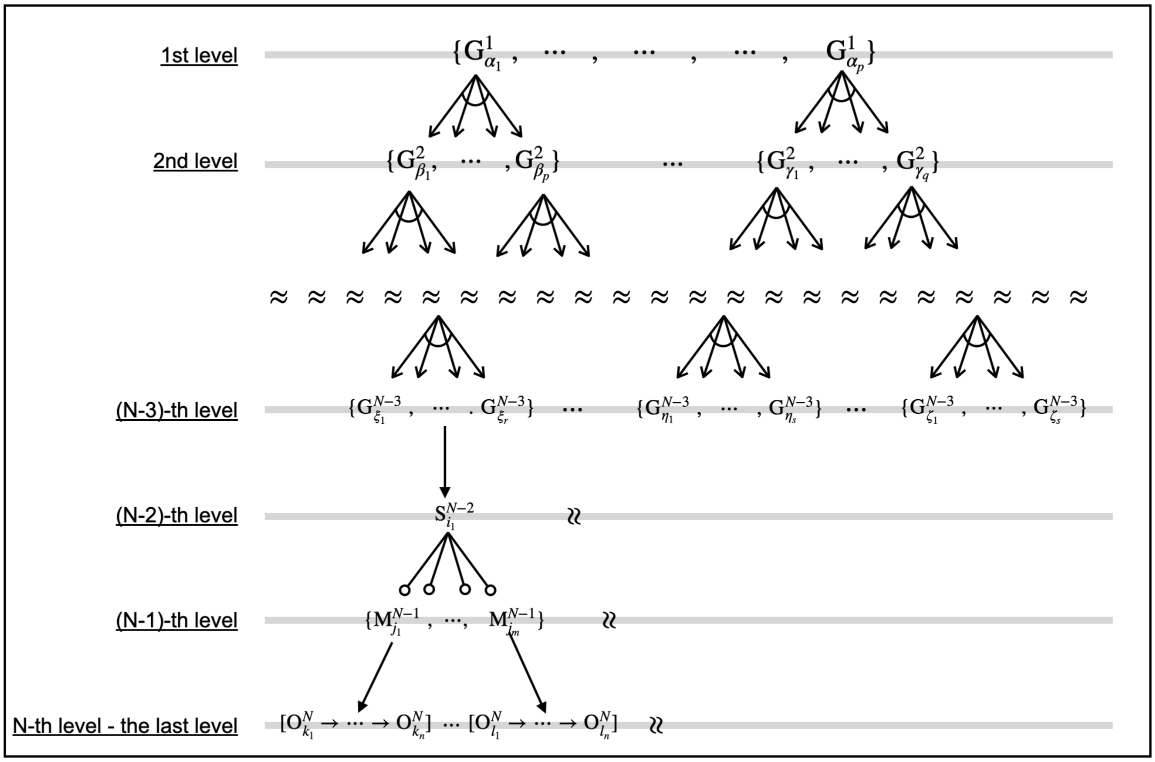


Figure 5. GOMS connection structure

world [17].

IV. Discussion: Deepening the Understanding of Behavioral Ecology

A. GOMS-construct Structure and Behavioral Ecology

1) GOMS-construct Structure: Figure 5 shows the overall picture of GOMS-construct, which has been constructed by experience, in a general form. The GOMS-construct is explained in the following by starting from the operator sequence, $\mathbf{O} = [\mathrm{O}_{k_1}^N \to \cdots \to \mathrm{O}_{k_n}^N]$, which is expanded at the lowest N-th layer, and working upward. At the immediately above the $(N-1)$-th layer, the method that connects to this operator sequence, $\mathbf{M} = \mathrm{M}_{j_1}^{N-1}$, exists. This can be regarded as a node that holds a pointer to this operator sequence, $\mathbf{O}$. This method serves to connect the goal, $\mathbf{G_i} = \mathrm{G}_{\xi_i}^{N-3}$, which resides two layers above it, at the $(N-3)$-th layer, with the operator sequence, $\mathbf{O}$, to achieve it. If there are multiple methods, $\{\mathrm{M}_{j_1}^{N-1}, \cdots, \mathrm{M}_{j_m}^{N-1}\}$, that can achieve $\mathbf{G_i}$, the selection rule, $\mathrm{S}_{i_1}^{N-2}$, is placed between these layers at the $(N-2)$-th layer.

Above the $(N-3)$-th layer, a hierarchical structure of goals is developed. The goals located at the top level, the first layer, $\mathbf{G_i^1} = \mathrm{G}_{\alpha_i}^1$, are expanded into a set of goals, $\mathbf{G^2} = \{\mathrm{G}_{\beta_1}^2, \cdots, \mathrm{G}_{\beta_p}^2\}$, at the second layer, and $\mathbf{G_i^1}$ is achieved by the achievement of all goals contained in $\mathbf{G^2}$. Hereafter, all goals at the first layer are expanded while maintaining this structure until the $(N-3)$-th layer. Note that in Figure 5, for convenience, the top-level goal is placed at the first layer and the N-th layer is represented as the lowest operator layer, but the depth of the hierarchy varies depending on the content of the top-level goal. Therefore, the concrete value of N varies depending on $\mathbf{G_i^1}$.

It would be appropriate to regard the individual G, O, M, S shown in Figure 5 as nodes that hold pointers connecting them to specific parts of the A-, B-, and C-memes. In the real behavioral situations, efficient use of memory is requited for smooth operation of the PCM process. Therefore, it is necessary to select and execute appropriate actions without continuously referring to the Perceptual-Multi-Dimensional Memory Frame, which is activated by P-resonance with the environmental information, while keeping in sync with the environment where the situation changes from moment to moment. Given this, the assumption that "there is an upper bound on the total number of G-, O-, M-, S-nodes available in C-resonance mediated by the GOMS structure" seems reasonable. The total number of goals is denoted as $\hat{G}$, the total number of methods as $\hat{M}$, the total number of selection rules as $\hat{S}$, the total number of operators as $\hat{O}$, the average depth of the hierarchy as $\bar{N}$, and the upper bound on the number of nodes as $\hat{C}$, which is a constant value. We then consider variations in the overall picture of GOMS-construct constructed through experience under the condition, $\hat{G} + \hat{O} + \hat{M} + \hat{S} \leq \hat{C}$, based on the relationship between each of the upper bounds.

An operator is an elemental perceptual, motor, or cognitive action; the execution of which produces a distinguishable change in the actor and/or the environment. Since the operator

Copyright (c) IARIA, 2025. ISBN: 978-1-68558-260-9

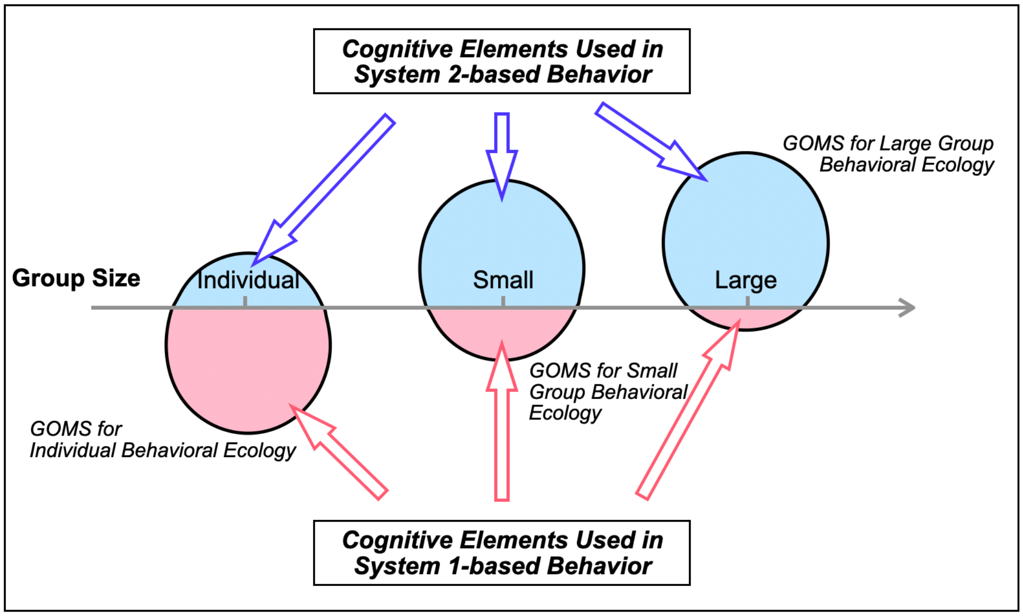


Figure 6. Changes in the Two Minds balance of GOMS components due to differences in behavioral ecology under brain processing capacity (time) constraints

is an elemental part of the construction of the method, $\hat{O}$ is presumed to be much smaller than $\hat{C}$. Then, how are the non-operator available nodes used? The top-level goals are expanded to sub-goals, eventually leading to the determination of a set of achievable methods to be defined under each goal. A method is a kind of goal that can be executed by the operators, pointing to a chunk of the operator sequence connected to it, so that the elements of the set of operators can be used as material to achieve the lowest goal that has been developed. If there are multiple methods that can achieve the goal, one method is selected based on the selection rule that defines the conditions for application of the method.

2) Characteristics of the GOMS-construct Structure: For an event $E(T)$ that occurs at time T, MHP/RT deals with it in its four processing modes [6]. They are as follows:

- System-2-Before-Event-Mode, which consciously considers $E(T)$ beforehand,
- System-1-Before-Event-Mode, which unconsciously adjusts its behavior to the environmental context immediately before $E(T)$,
- System-1-After-Event-Mode, which unconsciously adjusts the connections within the relevant Perceptual-, Behavior-, and Motor-Multi-Dimensional Memory Frame immediately after $E(T)$, and
- System-2-After-Event-Mode, which consciously reflects on $E(T)$ afterwards to adjust the connections within the Relation- and Word-Multi-Dimensional Memory Frame.

The GOMS-construct that each individual has developed should reflect the results of action selection in the System-1-After-Event-Mode and System-2-After-Event-Mode using the A-, B-, and C-memes in the System-2-Before-Event-Mode and System-1-Before-Event-Mode. By allocating more resources, i.e., brain processing power, to System-2-After-Event-Mode, he or she can construct a richer goal structure, which allows System-2-Before-Event-Mode to devote more resources to making accurate and reliable predictions in a variety of future situations he or she encounters. On the other hand, a sequence of methods involving successively occurring events, $E(T), \cdots, E(T{+}n)$, can be integrated into a single method by allocating more resources to System-1-After-Event-Mode. The integrated specialized method generates a specialized operator sequence for the corresponding sequence of events. In facing a variety of situations, the number of specialized methods will increase. Due to the limited processing capacity of the brain, either System-2-After-Event-Mode or System-1-After-Event-Mode will become dominant. Therefore, the following is predicted concerning the shape of GOMS-construct:

- If System-2-After-Event-Mode is dominant, then a goal-rich GOMS-construct, $\hat{G} \gg \hat{M}$, will be constructed.
- If System-1-After-Event-Mode is dominant, then a method-rich GOMS-construct, $\hat{G} \ll \hat{M}$, will be constructed.

3) Relationship between the Number of GOMS Components and the Balance of Conscious/Unconscious Processing: Figure 6 shows that the balance between System-1-After-Event-Mode-dominance and System-2-After-Event-Mode-dominance changes depending on the range of communities that the individual is directly and indirectly involved in during his/her life. In a behavioral ecology where individuals rarely interact with others, each individual can lead a sufficiently problem-free life by having a set of methods that are specific to the situations he or she encounters. Therefore, the relation, $\hat{M} \gg \hat{G}$, holds. As shown in the left portion of Figure 6, most actions are generated through unconscious execution of methods by System 1.

In the case of community-based living, each individual is expected to act according to the way he or she functions within the group he or she belongs to. When communication among group members is established in surface language, individuals are unable to perform elaborate inferences. Therefore, the relationship, $\hat{G} > \hat{M}$, is established, which is shown in the middle portion of Figure 6.

When a group belongs directly to a community and that community constitutes a society, i.e., the group belongs indirectly to the society, and/or when communication is done in structural language, the behavioral ecology becomes System-2-After-Event-Mode-dominant and the relationship, $\hat{G} \gg \hat{M}$, is established as shown in the right portion of Figure 6. The individuals can respond to various situations flexibly by allocating resources to the execution of System-2-Before-Event-Mode with careful use of the well-developed goal structure.

Figure 6 also shows the change in the GOMS-construct as the size of the group changes from individual, to small group, to large group. The number of elements that make up GOMS, $\hat{C}$, is limited by the constraints imposed on the processing capacities of brain. As the social relationships increase, the number of methods, which are System 1 elements, decreases through the reorganization of the goal structure by abstracting multiple individualized methods together. The elements, which have been used for System 1, are used by the System 2 elements. Meanwhile, the number of System 2 elements increases as the complexity of the relationship increases. In other words, by shifting to a behavioral ecology in which System 2 elements are more important than System 1 elements, the composition of elements in the entire GOMS will change to a composition with a rich goal structure that allows for more logical thinking.

B. GOMS and Meme

1) Mutual Development of GOMS and Meme: The existence of memes is a prerequisite for the generation of GOMS. GOMS also plays an important role in efficient action generation. The generated actions update Multi-Dimensional Memory Frame and contribute to meme formation. Thus, GOMS and memes are in a mutually developing relationship. Actions are generated in two ways: driven by System 1 or driven by System 2. The bias, i.e., the dominance of System 1 or System 2, in the generation of action, should affect the aspect of mutual development. This point is discussed in the following.

Figure 3 shows three types of memes. These memes are maintained through generations with imitation as the basic mechanism. A- and B-memes involve physical behaviors that are executed by connecting the Perceptual-, Behavior-, and Motor-Multi-Dimensional Memory Frame. Since A-memes are elemental in generating behavior and B-memes are combinations of elements of A-memes, they are different in granularity and do not mix with each other. The content of the inherited memes is almost invariant, since the content of physical behavior does not change significantly over time. The C-memes, on the other hand, are disconnected from physical behavior. It includes language activities with linguistic symbols and inference through the application of rules. The Word- and Relation-Multi-Dimensional Memory Frame are used for these activities. Linguistic symbols and rules are gradually updated under the influence of the social and natural environment surrounding each generation. A-, B-, and C-memes exist in parallel, without mixing with each other, and each is inherited from generation to generation.

GOMS covers orthogonally to the parallel meme structure and allows A-, B-, and C-meme elements to be combined with each other to efficiently generate effective actions in response to the real-world situations. This is accomplished by combining the elements in Multi-Dimensional Memory Frame in the form of a GOMS-construct. Since words are typical of memes, we can regard memes as carriers of meaning, i.e., *semantics*. GOMS can be thought of as *syntax* because it specifies how words are combined together.

C-memes represent inherited cultures. Cultures are diverse. Based on the discussion in the previous section, Section IV-A3, we can broadly distinguish between cultures that are rich in the goal structure of GOMS, G-culture, and cultures that are rich in the variety of methods, M-culture. Individuals acting in each culture acquire and act upon the inherited memes of that culture. The memes in the G-culture might be updated through System-2-After-Event-Mode, whereas those in the M-culture might be updated through System-1-After-Event-Mode. In either case, if the meme is deemed valid within the population in the updated structure, it will trigger a meme update. The update of a meme requires time for validation. Thus, it does not mean that the meme will be updated immediately.

Since a GOMS-construct links goals and operators, it guarantees corporeality for the goals present in the Word- and Relation-Multi-Dimensional Memory Frame. This ensures that even in the G-culture, the development of GOMS for various goals does not dissociate them from the real world. In other words, the connection of the Word- and Relation-Multi-Dimensional Memory Frame, to which G belongs, with Behavior- and Motor-Multi-Dimensional Memory Frame, to which M and O belong, guarantees the corporeality of the goal, G. By applying the GOMS-construct to memes, it is possible to make the meme, which is not linked to the real world as it stands, not free from the real world.

2) Common Understanding of Words and GOMS: Words are a typical example of memes [9] and the elements of C-memes. Words are the primary communication medium and are passed on from generation to generation [3]. Individuals make sense of words and understand the situation by referring to the context in which the words have been uttered. However, individual members of a community that share a C-meme may not assign a common meaning to a particular word, even when placed in a common context [12].

It is said that the number of words known by native English-speaking adults is 20,000∼30,000, and the number of words used in daily conversation is 3,000∼4,000. Conceptually known words are inherited as the elements of C-memes. However, the words used in daily activities unite the C-meme with the B- and A-memes, which are associated with corporeality, by means of GOMS. The goals in GOMS represented by symbols belonging to the C-memes are developed into the operators of GOMS belonging to the A-memes, and the meaning of goals can be shared as the B-memes as a sequence of operators, i.e., the methods of GOMS, that can be superficially observed as they carry out their daily activities.

3) Number of Top Located Goals and Behavioral Ecology: Culture and customs are examples of C-memes. Among them, what is desired to be achieved defines the goal structure of the community that inherits the C-memes. At the top level of the goal structure is the happiness goals [18] that members of a community commonly seek to achieve. Behavioral goals that lead to the achievement of the happiness goals exist beneath them [19].

In recent years, the characteristics of societies that rely on strong kinship relations and those in which individualism is prevalent have been discussed [20]. Based on our study, we can provide a perspective for understanding the behavioral ecology that forms in these societies. In the societies where the C-memes reflecting strong kinship are inherited, the number of goals that can exist is limited, and the System 1-driven behavioral ecology is formed as shown in the left portion of Figure 6. In contrast, in the societies with advanced individualism, each individual constructs his/her own goal structure, thus forming the System 2-driven behavioral ecology as shown in the middle and left portion of Figure 6.

In the latter case, many elements are used to construct the goal structure, and flexible action selection is achieved by flexibly replacing the higher-level goals depending on the situation. Since the replacement of the topmost happiness goal also occurs, the behavior is executed by switching between GOMS structures that are quasi-independent of each other. The fact that switching of the GOMS structure occurs can be described as a manifestation of the modalization of behavior. Modalization of behavior results in the appearance that an individual switches his or her behavioral norms depending on the situation.

This also does not necessarily guarantee that even if the same operator sequence is observed, the goal structure developing on top of that sequence is unique. And the possibility of misunderstandings, i.e., communication errors, arising from this cannot be excluded. The problems inherent in an individualistic society will appear here.

V. Conclusion

MHP/RT consists of the perceptual, cognitive (Two Minds), and motor processes that operate synchronously with changes in the environment, and the Multi-Dimensional Memory Frame that is used during action selection and execution of the PCM processes and is updated after action execution. The latter is an internal process of Multi-Dimensional Memory Frame, so it is executed separately from the PCM process. On the other hand, for the former, it is necessary to realize the connections between the perceptual process and Perceptual-Multi-Dimensional Memory Frame, and between the cognitive process and Word-, Relation-, and Behavior-Multi-Dimensional Memory Frame. MHP/RT introduces P-resonance for perception and C-resonance for cognition; the connections are realized by *resonance* shown in Figure 2. For P-resonance, we have introduced the basic senses as the mechanism in the preceding paper [4].

In this study, the mechanism of C-resonance was examined. In the Multi-Dimensional Memory Frame, there are memes, which are A-, B-, and C-memes, structured by three hierarchies as shown in Figure 3. These memes are mapped to each memory in the Multi-Dimensional Memory Frame and are linked to each other by sharing Perceptual-Multi-Dimensional Memory Frame as shown in Figure 2. For this reason, reality is ensured by perceptual information. On the other hand, C-resonance works under a time-constrained situation in which the PCM process must select and execute actions while synchronizing with changes in the environment, and connects the Multi-Dimensional Memory Frame and cognitive processes. In this study, we introduced GOMS proposed by Card, Moran, and Newell [5] as the mechanism to link Word-, Relation-, Behavior-, and Motor-Multi-Dimensional Memory Frame directly, without going through the Perceptual-Multi-Dimensional Memory Frame.

Each element of GOMS is represented as a node in the brain. The finite number of nodes that can be maintained allows different behavioral ecologies to emerge depending on how the number of nodes allocated to goals and selection rules operated by System 2 and the number of nodes allocated to methods and operators operated by System 1 are balanced. Regarding the C-memes, we examined the characteristics of behavioral ecology in the societies characterized by strong kinship, which inherit simple goal structures, and in the societies with a strong individualistic flavor, which inherit complex goal structures. Although the former is not expected to be flexible in action selection, it can achieve effective and efficient action selection and execution in stable social situations. In the latter, on the other hand, a modalized goal structure is maintained to cope with various situations, and the individual flexibly switches to the appropriate goal structure while selecting and executing actions. We also pointed out that although actions are observed as operator sequences, they are prone to communication errors caused by the non-uniqueness of the goal structure that develops on top of them.

The memes determine the content of the action. The PCM processes determine how to act. The GOMS structure intersects them. By viewing behavioral ecology from the perspective of GOMS, this study showed that static memes can be implemented and brought to life in behavioral ecology. Behavioral ecology is created by living organisms. On the basis of MHP/RT, the manifestation of memes in behavioral ecology has been clarified in the case studies [21][22]. This study is positioned as a proposal for a method to give life to static digital information while building on the results of these case studies.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 19K12246 / 19K12232 / 20H04290 / 22K12284 / 23K11334 , and National University Management Reform Promotion Project.

REFERENCES

[1] R. Dawkins, The Selfish Gene. New York City: Oxford University Press, 1976.
[2] R. Aunger, Ed., Darwinizing Culture: The Status of Memetics As a Science. Oxford Univ Pr on Demand, 1 2001.
[3] M. Kitajima, M. Toyota, and J. Dinet, "How Resonance Works for Development and Propagation of Memes," International Journal on Advances in Systems and Measurements, vol. 14, 2021, pp. 148–161.
[4] M. Kitajima et al., "Basic Senses and Their Implications for Immersive Virtual Reality Design," in AIVR 2024 : The First International Conference on Artificial Intelligence and Immersive Virtual Reality, 2024, pp. 31–38.
[5] S. K. Card, T. P. Moran, and A. Newell, The Psychology of Human-Computer Interaction. Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.
[6] 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.
[7] M. Kitajima, Memory and Action Selection in Human-Machine Interaction. Wiley-ISTE, 2016.
[8] 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.
[9] D. C. Dennett, From Bacteria to Bach and Back: The Evolution of Minds. W W Norton & Co Inc, 2 2018.
[10] S. Blackmore, The Meme Machine. OUP Oxford, 2000.
[11] F. J. Odling-Smee, K. N. Lala, and M. Feldman, Niche Construction: The Neglected Process in Evolution: The Neglected Process in Evolution. Princeton University Press, 2 2013.
[12] M. Kitajima et al., "Language and Image in Behavioral Ecology," in COGNITIVE 2022 : The Fourteenth International Conference on Advanced Cognitive Technologies and Applications, 2022, pp. 1–10.
[13] H. A. Simon, The Sciences of the Artificial, 3rd ed. Cambridge, MA: The MIT Press, 1996.
[14] J. R. Anderson, Rules of the Mind. Hillsdale, NJ: Lawrence Erlbaum Associates, 1993.
[15] J. R. Anderson and C. Lebiere, The Atomic Components of Thought. Mahwah, NJ: Lawrence Erlbaum Associates, 1998.
[16] J. R. Anderson, How can the Human Mind Occur in the Physical Universe? New York, NY: Oxford University Press, 2007.
[17] M. Kitajima, J. Dinet, and M. Toyota, "Multimodal Interactions Viewed as Dual Process on Multi-Dimensional Memory Frames under Weak Synchronization," in COGNITIVE 2019 : The Eleventh International Conference on Advanced Cognitive Technologies and Applications, 2019, pp. 44–51.
[18] D. Morris, The nature of happiness. London: Little Books Ltd., 2006.
[19] M. Kitajima, M. Toyota, and J. Dinet, "Guidelines for Designing Interactions Between Autonomous Artificial Systems and Human Beings to Achieve Sustainable Development Goals," International Journal on Advances in Intelligent Systems, vol. 15, 2022, pp. 188–200.
[20] J. P. Henrich, The weirdest people in the world: How the west became psychologically peculiar and particularly prosperous. Picador: Farrar, Straus and Giroux, 2021.
[21] M. Kitajima, M. Toyota, and J. Dinet, "Art and Brain with Kazuo Takiguchi - Revealing the Meme Structure from the Process of Creating Traditional Crafts -," in COGNITIVE 2023 : The Fifteenth International Conference on Advanced Cognitive Technologies and Applications, 2023, pp. 1–10.
[22] K. T. Nakahira, M. Kitajima, and M. Toyota, "Understanding Practice Stages for a Proficient Piano Player to Complete a Piece: Focusing on the Interplay Between Conscious and Unconscious Processes," International Journal on Advances in Life Sciences, vol. 16, 2024, pp. 164–177.

 ISBN: 978-1-68558-260-9

Neutralized Synchronic and Diachronic Potentiality for Interpreting Multi-Layered Neural Networks

Ryotaro Kamimura
Tokai University
Kitakaname, Hiratsuka, Kanagawa, Japan
e-mail: ryotarokami@gmail.com

***Abstract*—The present paper aims to propose a new method, called "di-synchronic potentiality," to unify or neutralize diachronic and synchronic potentialities in order to seek for the simplest form of a network, referred to as the prototype. This method is necessary because the prototype is deeply hidden within surface networks, making it challenging to detect. The synchronic potentiality is measured by the complexity of connection weights at a specific learning time, while the diachronic potentiality is time-dependent. These potentialities tend to decrease as a natural property of learning. While this reduction is effective in eliminating unnecessary weights, it may also lead to the eventual elimination of important weights. The di-synchronic potentiality aims to unify or neutralize the reduction forces of diachronic and synchronic potentialities to increase synchronic potentiality. With this method, the synchronic potentiality does not necessarily increase, but at the very least, the reductive force is weakened or neutralized for solving the collision between the two types of reduction forces. The method was applied to artificial data simulating the bankruptcy of companies, with both linear and nonlinear relations between inputs and targets. The results confirmed that there were strong and repeated forces striving to obtain the simplest form of potentiality. Ultimately, the method successfully produced representations with improved generalization, while simultaneously achieving the simplest relations between inputs and outputs for easier interpretation. From these results, we can conclude that detecting the simplest prototype can eventually lead to discovering more complex yet interpretable relations between inputs and outputs.**

Keywords-diachronic; synchronic; di-synchronic; interpretation; potentiality; prototype

I. Introduction

A. Simplification Forces

The present paper aims to demonstrate the existence of simplification forces in neural networks. These forces can be observed through the simplest network, called the "prototype network," which is deeply hidden within a vast number of different surface networks. Naturally, as the complexity of neural networks and input patterns increases, this simplification hypothesis might seem irrelevant for practical purposes. However, neural networks attempt to replicate only a small fraction of cognitive functions. Compared with the human brain, the primary learning mechanisms in neural networks are expected to be simpler and more straightforward than we might imagine. This implies that even as the complexity of modern neural networks and their input patterns appears to grow, the fundamental inference mechanism remains inherently simple.

Furthermore, the existence of simplification forces is practically useful for achieving simplified interpretations. Neural networks have been applied across many fields, but a major challenge is understanding the inference mechanisms behind these complex systems. If simplification forces exist, the key task in interpreting neural networks becomes understanding the simplified networks resulting from these forces. Naturally, more complex features beyond the simplified ones may be necessary for addressing practical problems. However, if we assume that even in complex networks, much simpler prototype networks operate at the deepest level, interpretation becomes significantly easier. This suggests that by understanding and referencing the simplest network, we can better grasp why seemingly complex features are required in actual processing.

B. Neutralized Potentiality Reduction

To simplify multi-layered neural networks, we have proposed the potentiality method to control the number of absolute connection weights. The potentiality is similar to conventional entropy, and as is well known, learning can be represented by entropy reduction [1]. Following this entropy hypothesis, we can assume that learning is a process of reducing the potentiality of networks. One of the major problems is that potentiality reduction focuses solely on decreasing the strength of connection weights, which does not necessarily eliminate unimportant weights. Instead, it may also reduce important weights. Thus, it is crucial to control this reduction process to preserve important information.

To address the issue of exclusive potentiality reduction, we introduce two types of potentiality: "diachronic" and "synchronic" potentiality, and propose a method to neutralize or unify them to weaken the force of potentiality reduction. The synchronic potentiality is equivalent to the conventional potentiality or entropy defined without considering time-dependent changes during the learning process. In contrast, diachronic potentiality represents the time-dependent learning process. Like synchronic potentiality, diachronic potentiality also aims to reduce the corresponding potentiality. To weaken the forces of synchronic and diachronic potentiality reduction, we embed the diachronic potentiality into the synchronic potentiality. This approach is referred to as diachronized synchronic potentiality or "di-synchronic potentiality." The control of di-synchronic potentiality seeks to simultaneously reduce information and maximize it, or at least weaken the force of potentiality reduction through the interplay between these two

 ISBN: 978-1-68558-260-9

types of potentiality reduction. Similar to the hypothesis that important information should be maximized while unnecessary information should be minimized, this di-synchronic potentiality control incorporates the effect of potentiality maximization into the process of potentiality minimization.

C. Detecting the Prototype Network

This di-synchronic potentiality control can be used to detect the simplest form of a network, or the prototype network, deeply hidden within seemingly complicated surface neural networks. In this paper, we do not extract the prototype network directly from input patterns; rather, the prototype is expected to be self-organized if all necessary network components are provided before learning. We assume that behind any multi-layered neural network, there exists the simplest network without hidden layers, where all components are linearly and separately connected. These properties are not derived from receiving input patterns but are ideally self-organized, independently of any input patterns. However, in practice, it is impossible to achieve full self-organization automatically. Thus, it is more accurate and moderate to state that the prototype can only be self-organized by processing some input patterns initially. The prototype is self-organized by incorporating information necessary to conform to the norms inherent to the prototype. The di-synchronic potentiality is introduced to extract this prototype, which is deeply hidden within seemingly complex network configurations and input patterns. We aim to control the potentiality until we uncover the simplest form of the prototype, even at significant cost.

D. Main Contributions

Then, the outline and the main contributions of this paper can be summarized as follows:

- The present paper aims to clarify the simplification forces hidden in multi-layered neural networks, which are assumed to be represented in terms of the simplest prototype networks.
- These prototypes are presumed to be deeply hidden within surface neural networks, and it is necessary to uncover them using specific methods. We have proposed the potentiality method to reduce and simplify the configuration of connection weights. However, it is challenging to preserve important connection weights through potentiality reduction alone. In this context, we propose an additional potentiality method to increase the potentiality or at least to weaken the reduction forces, thereby controlling the process of potentiality reduction.
- The new method is realized by combining the conventional synchronic potentiality with diachronic potentiality, resulting in a new type of potentiality called “di-synchronic potentiality.” This di-synchronic potentiality is achieved by deploying synchronic potentiality over diachronic processes or learning processes. While this potentiality does not necessarily increase the synchronic potentiality, it can at least weaken the reduction forces of synchronic potentiality.

E. Paper Organization

In Section 2, we summarize several related works, such as internal, prototype, interpretable, distilling, mutual information simplification. In Section 3, we present how to compute the synchronic and diachronic potentiality, and the di-synchronic potentiality. In Section 4, we applied the method to an artificial data set designed to simulate a business data set, incorporating both linear and non-linear relations between inputs and targets. The results showed that the reduction of synchronic potentiality was weakened, preventing the network from excessively reducing the potentiality. The di-synchronic potentiality repeatedly and intensely sought to extract the prototype during the learning process. Connection weights became closer to those of the assumed prototype, and improved generalization was observed. Ultimately, the results confirmed that as interpretation became easier, generalization performance also improved.

II. Related Work

A. Internal and External Simplification

As mentioned in the introductory section, one of the major problems in neural networks is the inability to interpret the inference mechanism. To address this issue, many different types of interpretation methods have been developed, all of which are directly related to simplifying network configurations and input patterns. To explain the main characteristics of our method in comparison to conventional simplification methods, we introduce two types of simplification: external and internal simplification. External simplification aims to simplify the network configuration depending on inputs and, more directly, to simplify input patterns themselves. In contrast, internal simplification focuses not on input patterns but on the network itself. Ideally, the network should be self-organized without external stimuli. However, self-organization without any external stimuli is impossible, so this simplification is moderated to allow networks to self-organize with some initial external stimuli. We explain four simplification procedures in comparison to our method, namely, prototype simplification, interpretable simplification, distilling simplification, and mutual information-theoretic simplification.

B. Prototype Simplification

The prototype approach has been one of the major simplification procedures in neural networks. Here, we use the prototype to describe the simplest form achievable through the simplification forces inherent in the network. Prototype-based approaches have been studied extensively since the early stages of learning, under names, such as vector quantization, competitive learning, and self-organizing maps [2]. These methods aim to identify a small number of representative vectors for many input patterns. In recent deep learning research, prototype learning has gained attention [3], as the volume of input patterns to be processed has grown larger and increasingly heterogeneous. Simplifying these complex and heterogeneous patterns has become urgent, giving rise to methods like one-shot and few-shot learning to represent many inputs using a few representative patterns. Similarly, zero-shot learning [4] has been developed to incorporate the abstract and semantic properties of input patterns. These methods exemplify external simplification, representing a large number of input patterns with a smaller set of representative inputs and more abstract features. In contrast, our approach focuses on the network itself, aiming to determine which network configuration should be organized when all network components are given before learning.

C. Interpretable Simplification

Second, most interpretation methods can be classified as external simplification. Among these, one prevalent approach is localized interpretation, which focuses on specific instances rather than the overall inference mechanism. Unlike global interpretation methods [5]–[7], localized interpretation has proven effective in practical applications due to its simplicity and the urgent need for explainability. Linear and local simplifications have been widely used in interpretation methods, replacing complex non-linear models with corresponding linear models [8]–[10]. These simplified local models aim to interpret surface networks, which are diverse and heterogeneous. However, our approach seeks to uncover the prototype hidden deeply within surface models. Surface models are produced through transformational rules and may be too complex to understand, but the underlying prototype is expected to be much easier to interpret.

D. Distilling Simplification

Third, network compression simplifies neural networks by replacing complex networks with smaller ones [11]–[13]. This type of simplification has become increasingly necessary as networks grow larger to process more input patterns. However, interpreting how all components in such networks operate to produce outputs is a significant challenge. Most compression methods represent typical external simplification, replacing the original network with smaller, unrelated networks. These methods attempt to interpret the smaller "student" networks under the assumption that their inference mechanism is similar to the original network's. However, this assumption may not always hold true. In contrast, the internal simplification we propose compresses the original multi-layered neural network directly, retaining the original network's information. Thus, interpreting and explaining the compressed networks is more directly connected to the original multi-layered networks.

E. Mutual Information-Theoretic Simplification

Fourth, information-theoretic methods also relate to network simplification, albeit in more abstract ways. Multiple network configurations can be represented by more simple and abstract information content, and the objective of learning can be considered as necessary information acquisition. In particular, mutual information has played a significant role in neural networks. Notable examples include the well-known maximum mutual information preservation [14], which aims to retain as much relevant information as possible, and the information bottleneck method [15], which seeks to maximize necessary information while minimizing unnecessary information. Despite their utility, mutual information-based methods face challenges, such as computational complexity and ambiguous interpretations of the results. These issues arise because mutual information involves contradictory operations: entropy maximization and conditional entropy minimization. Recent information bottleneck methods introduce additional mutual information computations to balance compression and relevant information preservation [16], but computational complexity remains a challenge.

Our bi-synchronic potentiality method shares the goal of acquiring simplified, necessary information but differs in three key ways. First, it focuses on information necessary for self-configuring neural networks with minimal external input, deepening the level of simplification compared to conventional methods. Second, it embeds diachronic potentiality reduction into synchronic potentiality reduction, neutralizing and weakening potentiality reduction forces. Third, it aims to make network interpretation easier by transitioning from surface-level to prototype-level interpretation. Fundamentally, our method seeks to examine the information required to configure networks when all resources are provided before learning begins. Unlike conventional information-theoretic methods that optimize configurations to represent input patterns efficiently, our approach tries to create a framework to prepare for potential input patterns.

III. Theory and Computational Methods

A. Synchronic and Diachronic Potentiality

The prototype represents the state of network configuration in terms of connection weights. So far, we have defined the potentiality without considering time-dependent factors. This paper introduces a time-dependent potentiality, termed "diachronic potentiality," while the original potentiality, which does not account for time-dependent factors, is referred to as synchronic potentiality. The potentiality function was developed to simplify the conventional entropy function. By omitting the logarithmic function, potentiality becomes computationally efficient and easier to interpret. One major issue is that potentiality is designed to decrease its strength. Minimizing synchronic potentiality is expected to reduce redundant information and highlight important information. However, this reduction does not necessarily extract important information, as it focuses solely on reducing strength without adequately considering the properties of network components.

To address this issue, we introduce a factor that augments potentiality or at least weakens potentiality reduction to enhance the extraction of important information. For this purpose, we define diachronic potentiality to capture the time-dependent properties of learning. Diachronic potentiality characterizes the entire learning process and naturally tends to decrease as learning progresses, similar to synchronic potentiality. By embedding diachronic potentiality into synchronic potentiality, the reduction effect can be weakened through a process called neutralization. This combined potentiality

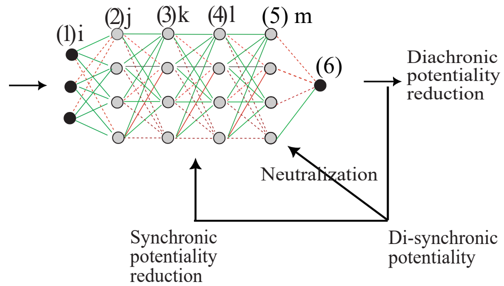


Figure 1. Synchronic, diachronic and di-synchronic potentiality by neutralization.

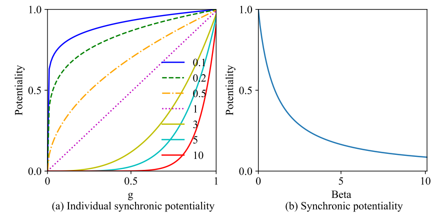


Figure 2. Individual synchronic potentiality (a) and synchronic potentiality as a function of the parameter β_{syn} (b). The potentialities are normalized for easy comparison.

is referred to as "di-synchronic potentiality," reflecting the integration of diachronic and synchronic aspects of potentiality reduction. Figure 1 illustrates the concept of this neutralization. In the figure, diachronic potentiality operates throughout all learning steps, while synchronic potentiality is effective only at specific steps of the overall learning process. Both types of potentiality aim to reduce their respective strengths. However, through neutralization, achieved by embedding diachronic potentiality into synchronic potentiality, the two can be unified. The concept of neutralization signifies an effort to increase synchronic potentiality or, when this is not feasible, to mitigate the strength of synchronic potentiality reduction.

B. Synchronic Potentiality Reduction

Let us begin with the definition of synchronic potentiality, as all other types of potentiality are based on this concept. The synchronic potentiality is defined using the absolute values of connection weights. For simplicity, we consider only one hidden layer, from the nth layer to the $n+1$th layer, denoted as $(n, n+1)$, as shown in Figure 1. The individual synchronic potentiality for $(n, n+1)$ is computed as:

$$u_{jk}^{(n,n+1)} = \mid w_{jk}^{(n,n+1)} \mid, \tag{1}$$

where all weights are assumed to be greater than zero for simplicity. By normalizing with the maximum value, we define the relative synchronic potentiality as:

$$g_{jk}^{(n,n+1)} = \gamma_{syn} \left[\frac{u_{jk}^{(n,n+1)}}{\max_{j'k'} u_{j'k'}^{(n,n+1)}} \right]^{\beta_{syn}}, \tag{2}$$

where the maximum operation is taken over all connection weights in the layer, and β_{syn} is a parameter controlling the strength of the potentiality. The parameter γ_{syn} is introduced for computational purposes, as this potentiality may excessively reduce the strength of connection weights. By summing all the relative potentialities, we obtain the final form of synchronic potentiality:

$$G^{(n,n+1)} = \gamma_{syn} \sum_{jk} \left[\frac{u_{jk}^{(n,n+1)}}{\max_{j'k'} u_{j'k'}^{(n,n+1)}} \right]^{\beta_{syn}}. \tag{3}$$

Figure 2(a) illustrates synchronic potentiality as a function of the strength of individual synchronic potentiality. As the parameter β_{syn} increases, the majority of connection weights are forced to decrease, thereby reducing the potentiality. Figure 2(b) shows synchronic potentiality as a function of the parameter β_{syn}. As explained above, synchronic potentiality decreases gradually with increasing parameter strength.

New weights at the $(t+1)$th learning step are obtained by multiplying the weights at the tth step by the corresponding potentiality:

$$w_{jk}^{(n,n+1)}(t+1) = \gamma_{syn} \left[\frac{u_{jk}^{(n,n+1)}}{\max_{j'k'} u_{j'k'}^{(n,n+1)}} \right]^{\beta_{syn}} w_{jk}^{(n,n+1)}(t). \tag{4}$$

In this learning rule, the individual potentiality is multiplied by the corresponding weight. The weights are incorporated to facilitate learning, as the strength of connection weights must be considered for effective learning. Learning can theoretically proceed with only the potentiality term, without actual weights, although the parameter γ_{syn} would need careful tuning for stable learning. In this paper, we adopt this formulation for computational convenience.

C. Diachronic Potentiality Reduction

The diachronic potentiality is defined by considering the entire sequence of learning steps. The individual diachronic potentiality at the tth learning step (epoch) is defined solely based on the time step t:

$$v_t = t. \tag{5}$$

The relative individual diachronic potentiality is then obtained by normalizing the potentiality with its maximum value:

$$h_t = \gamma_{dch} \left[\frac{v_t}{\max_{t'} v_{t'}} \right]^{\beta_{dch}}, \tag{6}$$

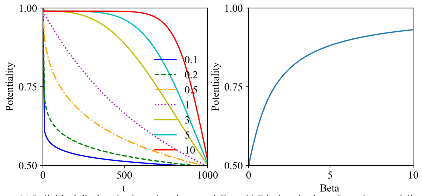
where γ_{dch} is a scaling parameter, and β_{dch} controls the strength of the diachronic potentiality. The total diachronic potentiality is defined as:

$$H = \gamma_{dch} \sum_{t} \left[\frac{v_t}{\max_{t'} v_{t'}} \right]^{\beta_{dch}}. \tag{7}$$

The final form of the weight update rule is given by:

$$w_{jk}^{(n,n+1)}(t+1) = \gamma_{dch} \left[\frac{v_t}{\max_{t'} v_{t'}} \right]^{\beta_{dch}} w_{jk}^{(n,n+1)}(t). \tag{8}$$

This learning equation indicates that the strength of the weights is reduced progressively over the course of learning.



(a) Individual diachronized synchronic potentiality (b) Diachronized synchoronic potentiality

Figure 3. Individual di-synchronic potentiality (a) and di-synchronic potentiality as a function of the parameter β_{dia} (b).

The effect of the diachronic potentiality increases as the learning step progresses. As the diachronic parameter β_{dch} increases, the diachronic potentiality is forced to become smaller, signifying the application of stronger potentiality reduction. The diachronic potentiality is defined in relation to the corresponding synchronic potentiality, and it tends to decrease as the parameter β_{dch} increases, as illustrated in Figure 2.

D. Di-Synchronic Potentiality Augmentation

We aim to neutralize the synchronic and diachronic potentiality into a new form, called the "di-synchronic potentiality." Specifically, we attempt to reduce the effects of synchronic and diachronic potentiality reduction to a certain extent. As mentioned above, this process is referred to as neutralization, where we balance the forces of potentiality reduction. One possible approach is to replace the synchronic parameter β_{syn} with the diachronic individual potentiality:

$$\text{neutral}_{jk}^{(n,n+1)}(t) = \gamma_{syn} \left[\frac{u_{jk}^{(n,n+1)}}{\max_{j'k'} u_{j'k'}^{(n,n+1)}} \right]^{h_t}, \qquad (9)$$

where h_t is the individual relative diachronic potentiality. This embedding is shown to be effective for neutralization.

Figure 3(a) illustrates the individual di-synchronic potentiality as a function of the number of learning steps for different values of the parameter β_{dia}. As shown in the figure, as the strength of the parameter increases, the strength of the individual potentiality also increases. Figure 3(b) presents the di-synchronic potentiality as a function of the parameter β_{dia}. It can be observed that the di-synchronic potentiality gradually increases as the parameter increases. This implies that while the diachronic potentiality decreases as the diachronic parameter increases, the di-synchronic potentiality can still be enhanced.

The learning procedure is finally expressed as:

$$w_{jk}^{(n,n+1)}(t+1) = \text{neutral}_{jk}^{(n,n+1)}(t)\; w_{jk}^{(n,n+1)}(t). \qquad (10)$$

By employing this procedure, we aim to augment the synchronic potentiality while neutralizing the effects of potentiality reduction.

Lastly, it is important to note that several additional potentialities are required to fully describe the learning process. Due to page limitations, explanations are provided as needed.

IV. Results and Discussion

A. Experiment Outline

The data set was created by imitating an actual business data set used to estimate the bankruptcy of companies [17]. The objective of the experiment is not to improve generalization but to interpret how bankruptcy occurs and what the major causes of bankruptcy are. In more technical terms, we aim to understand the relationships between inputs and outputs. In the actual data set, linear and non-linear relationships are naturally mixed, making it seemingly impossible to explicitly understand the relationships between inputs and outputs. To address this situation, we created a data set with both linear and non-linear relationships between inputs and targets artificially. This artificial data set allows us to explicitly understand how neural networks respond to specific inputs to produce outputs.

The number of input variables was seven. Of these, input No.6 (financial stability) and input No.7 (market sentiment) were created using exponential functions to stress the lower or higher values. Input No.5 was also created non-linearly using the sine function to represent large variations. Figure 4 shows the actual normalized correlation coefficients between inputs and outputs. Inputs No.1 to No.4 were created linearly, while the remaining inputs were created non-linearly, as described above. For example, input No.7 showed the lowest correlation coefficient among the inputs but was related non-linearly to the target. The problem is to examine whether the potentiality method can distinguish between these two types of inputs.

The number of input patterns was 1000, and the number of hidden layers was ten, with ten neurons in each hidden layer. We used the PyTorch program package, where almost all parameter values were set to default values to ensure easy reproduction of the results presented here. The experiment was designed to make the neural networks as close as possible to the prototype network, which is assumed to be hidden within the surface networks. The prototype network is the simplest form, and the connection weights in this paper were computed using the correlation coefficient between inputs and targets of the training data set. Next, we attempted to use the di-synchronic potentiality to neutralize the effect of synchronic potentiality reduction. The final multi-layered neural network was compressed into the simplest network without hidden layers. Compression was conducted layer by layer. We then compared the estimated and compressed networks with the assumed simplest network. Our objective is to demonstrate that the di-synchronic potentiality could be used to make the actual neural networks as close as possible to the prototype network.

Now, the main findings can be summarized as follows:

- Our method was able to weaken the forces of synchronic potentiality reduction. This means that the di-synchronic potentiality was effective in controlling the process of synchronic potentiality reduction.
- The results confirmed that the di-synchronic potentiality could regulate the synchronic potentiality to increase gener-

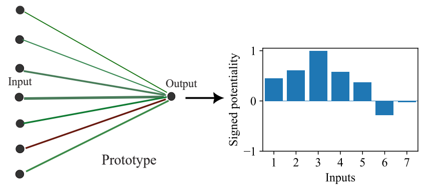


Figure 4. Supposed prototype (left) and the strength of connection weights (right).

alization performance by increasing the diachronic parameter β_{dia}.

- In addition, the di-synchronic potentiality repeatedly attempted to increase the ratio of potentiality, measuring similarity to the supposed prototype, even in the later stages of learning. This indicates that the potentiality could force networks to become closer to the prototype, while maintaining improved generalization.
- The experimental results confirmed that the di-synchronic potentiality could simplify a ten-layered neural network into the simplest prototype network with better generalization. We were able to simplify the internal representations created by the neural network, while still preserving improved generalization.

B. Potentiality and Generalization

The results confirmed that the di-synchronic potentiality, controlled by the diachronic parameter β_{dia}, could increase the synchronic potentiality, though moderately. This potentiality augmentation was associated with improved generalization.

Figure 5 shows the synchronic potentiality (left), entropy (middle), and testing and validation accuracy (right) as a function of the number of steps. When the linear activation function was used in Figure 5(a), both the potentiality and entropy took relatively higher values, but they remained almost unchanged, as the potentiality control was not introduced. When the parameter β_{dia} increased from 0 (b) to 1.2 (f), the synchronic potentiality (left) tended to have higher values. Since the differences were relatively small, we plotted these potentialities in one graph in Figure 6. It is clear that the values of synchronic potentiality increased gradually as the parameter β_{dia} increased. This indicates that, although the synchronic potentiality tended to decrease, a force to increase it could be observed as the parameter β_{dia} increased. In addition, the entropy (middle) exhibited this trend more clearly. When the parameter β_{dia} was small, the entropy remained higher in the early stages of learning and then decreased considerably. As the parameter increased further, the entropy tended to stay at higher values throughout the learning process.

The right figures in Figure 5 show testing and validation accuracy as a function of the number of steps. Generalization accuracy initially increased and then decreased rapidly as the parameter β_{dia} increased from 0 (b) to 0.4 (d). When the parameter was relatively small, the effect of synchronic potentiality tended to be stronger. As the parameter increased further, generalization accuracy gradually stabilized and began

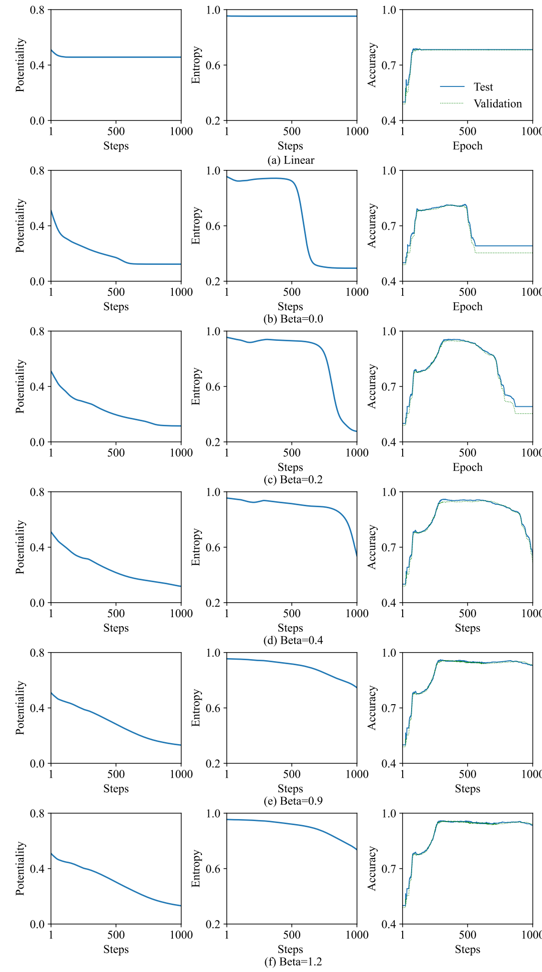


Figure 5. Synchronic potentiality (left), entropy (middle), and testing and validation accuracy (right) as a function of the number of steps (epochs), when the parameter β_{dia} increased from 0 (b) to 1.2 (f). In addition, results with the linear activation function was added (a).

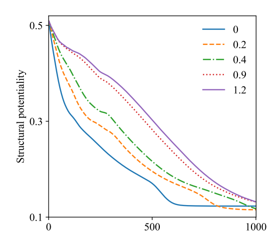


Figure 6. Synchronic potentialities as a function of the number of steps, when the parameter β_{dia} increased from 0 to 1.2.

to show relatively higher values. Finally, when the parameter reached 1.2 (f), the generalization accuracy was the highest.

The di-synchronic potentiality was shown to be effective in weakening the synchronic potentiality reduction, and this effect was clearly related to improved generalization.

C. Ratio Potentiality

Then, we examined to what extent the final networks could be close to the supposed prototype network. We used the ratio potentiality, representing the ratio of estimated potentiality to the supposed potentiality. As the ratio potentiality increases, the estimated prototype becomes closer to the supposed prototype. The results confirmed that the ratio potentiality showed higher values initially, and in particular, as the diachronic parameter increased, the force to achieve higher ratio values became stronger. Eventually, when the diachronic parameter increased to 1.2, which produced the best generalization, we could extract relatively higher ratio values three times during the learning process.

Figure 7 shows the ratio potentiality (left), divergence (middle), and correlation coefficients between the supposed and estimated prototypes (right). First, we observed that the correlation coefficients between the estimated and supposed prototypes were much higher, indicating close relations between the two prototypes. The ratio potentiality (left) remained almost flat throughout the entire learning process with the linear activation function (a). When the parameter β_{dia} was zero, the ratio potentiality tended to have higher values in the initial stages of learning, and then decreased. When the parameter β_{dia} increased to 0.2 (c) and 0.4 (d), only one peak appeared at the beginning of the learning process. When the parameter increased to 0.9 (e), two peaks in the ratio potentiality were observed. Finally, when the parameter reached 1.2 (f), three peaks were evident. This indicates that the di-synchronic potentiality forced the networks to resemble the prototype as much as possible. Comparing the results with those obtained from the divergence (middle) and correlation coefficients (right), the ratio potentiality more explicitly demonstrated the tendency to acquire the prototype.

The results confirmed that simplification forces were present in the neural networks, as evidenced by the higher ratio potentiality. Additionally, we observed many peaks in the ratio potentialities, meaning that these simplification forces were very strong throughout the entire learning process.

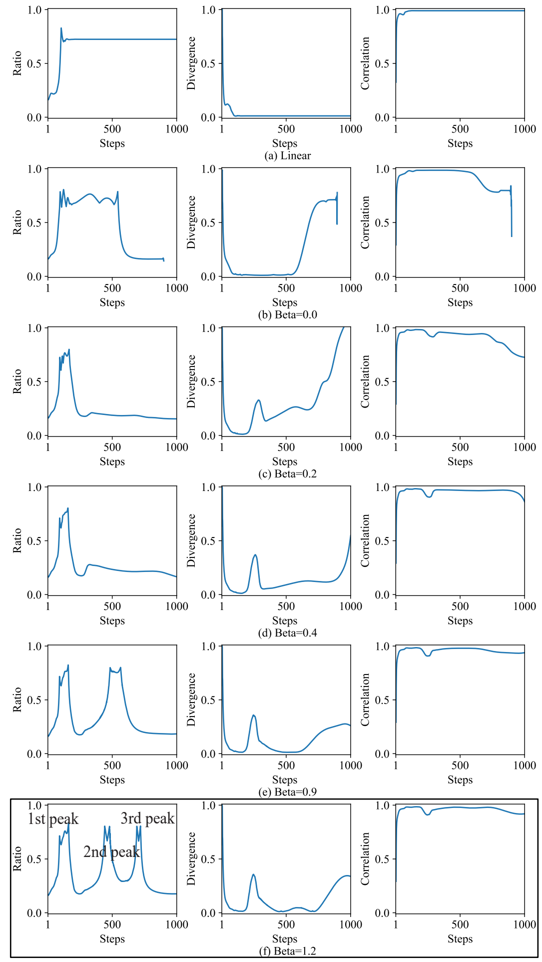


Figure 7. Ratio potentialities (left) and divergence (middle), and correlation coefficient (right), when the linear activation function was used (a), and the parameter β_{dia} increased from 0 (b) to 1.2 (f).

D. Interpretation of Rotating Individual Ratio Potentialities

1) Signed Individual Potentialities: The results confirmed that the signed individual potentialities, corresponding to the actual connection weights, were close to the prototype. Then, when we examined the ratio potentialities, the three ratio peaks exhibited higher ratio potentialities, indicating that the networks were close to the supposed prototype three times during the entire learning process.

Figure 8 (a) shows signed individual potentialities, namely, actual and normalized connection weights, when the parameter β_{dia} increased to 1.2, resulting in the best generalization. Comparing with the supposed prototype in Figure 4, the potentialities were quite close to the supposed prototype. However, gradually, the final input No.7 tended to become stronger. Figure 8 (b) shows individual ratio potentialities. For the three peaks, the potentialities were considerably high, indicating that the final connection weights were close to the prototype. However, in other cases, only input No.7 had overwhelmingly large values. The input No.7 was created with non-linear relations between inputs and outputs.

The results confirmed that there were strong simplification forces in the neural networks. By examining these individual potentialities, we observed that improved generalization could be obtained not only by detecting non-linear relations between inputs and outputs but also by clearly detecting linear relations.

2) Interpretation by Rotation of Individual Ratio Potentialities: We present the results by rotating the potentialities as the parameter β_{rat} of the ratio potentiality increases. As the parameter increased, the number of stronger potentialities became smaller. For the three peaks, the linear input No.4 and the non-linear inputs No.6 and No.7 were used to infer the outputs. The combination of linear and non-linear relations could be used to improve generalization.

For easier understanding of the ratio potentialities, we rotated the potentialities by changing the parameter β_{rat} for the ratio potentialities in Figure 9. When the parameter was 0.1 (a), almost all ratio potentialities became higher. As the parameter increased from 1 (b) to 5 (d), a progressively smaller number of potentialities remained relatively stronger. For the three peaks, input No.4, No.6, and No.7 became relatively higher and more important. Input No.4 was created with a linear relation, while the other two inputs were created with non-linear relations. Thus, the combination of linear and non-

Copyright (c) IARIA, 2025. ISBN: 978-1-68558-260-9

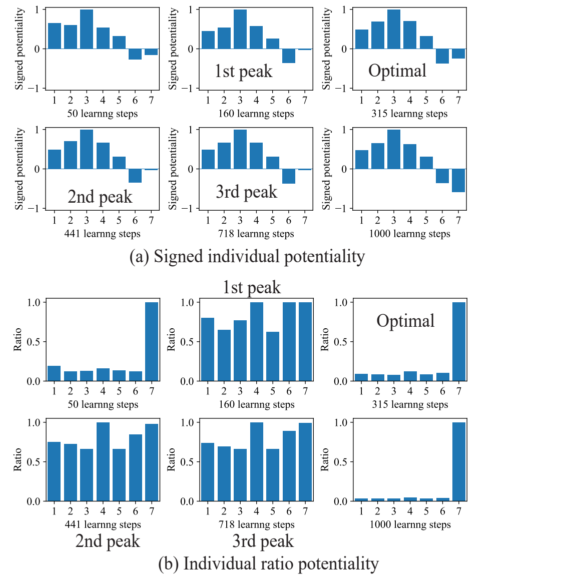


Figure 8. Individual ratio potentiates, when the parameter β was 1.2 with the best generalization.

linear relations should be used to improve generalization. However, for the three peaks, the ratio potentiality took very high values, and the relations between inputs and outputs could be interpreted easily by examining the linear correlation coefficients between inputs and outputs.

The results confirmed that the networks could show improved generalization based exclusively on the non-linear relation between inputs and outputs, but they could also produce approximately the same results with both linear and non-linear relations.

E. Prototype-based Interpretation

Finally, we should again examine the highest peaks in the ratio potentiality. The results confirmed that the networks could be close to the linear prototype, achieving improved generalization almost equivalent to the highest value by focusing on the non-linear relations. This means that the bi-synchronic potentiality could produce simplified and linear connection weights for easier interpretation with better generalization.

Table I shows the summary of the three peaks of ratio potentiality when the parameter was 1.2, with the best generalization. As shown in the table, the highest testing accuracy was 0.959 with 314 learning steps. However, the ratio potentiality was very low (0.224), and the divergence was larger (0.108). As explained in the previous section, this highest peak was obtained by considering exclusively input No.7 with non-linear relations.

For the first peak of ratio potentiality, with 159 learning steps, the ratio potentiality became higher (0.836), and the divergence was smaller (0.014). However, the testing accuracy was the lowest (0.783). This suggests that it seems impossible to increase generalization with the simple linear relations in the prototype network. However, for the second peak, the ratio potentiality was also quite large (0.805), and the divergence was the smallest (0.011). In particular, the testing accuracy

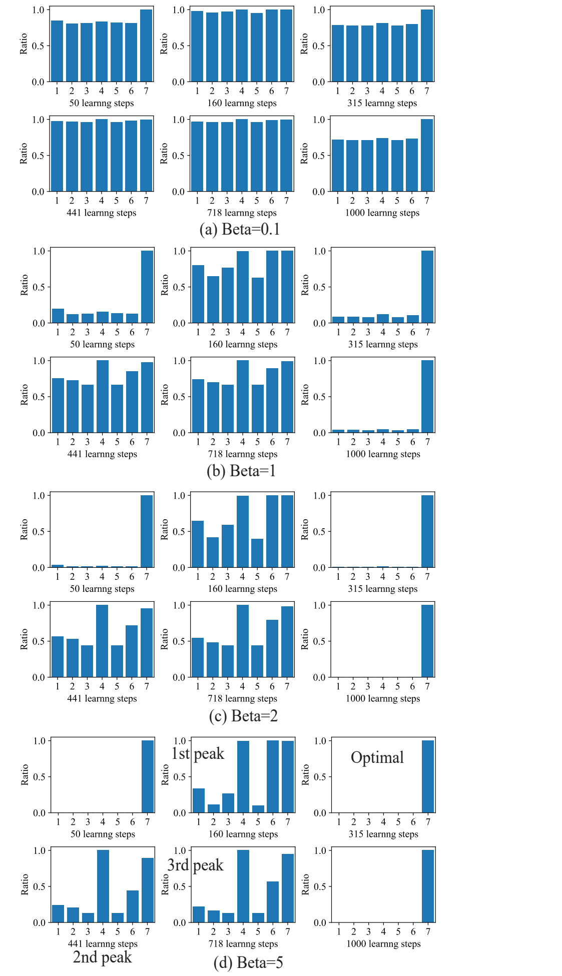


Figure 9. Rotating the ratio potentiality, when the parameter β_{rat} was increased from 0.1 (a) to 5 (d).

was 0.953, which was quite large, though smaller than that of the optimal case. For the third peak, with 717 steps, the ratio potentiality was still high (0.806), and the divergence was the second lowest (0.012). The generalization accuracy was 0.939, which was smaller than that of the optimal case, but sufficiently large.

This shows that the networks could be similar to the prototype network, whose simple relations between components make it possible to interpret the final results more easily, while still maintaining improved generalization.

TABLE I. TESTING ACCURACY, WHEN THE RATIO POTENTIALITY TOOK THE PEAKS WITH THE OPTIMAL CASE AND WITH THE BEST GENERALIZATION PERFORMANCE.

Peak	Step	Ratio Potent	Divergence	Testing
1st	159	0.836	0.014	0.783
2nd	440	0.805	0.011	0.953
3rd	717	0.806	0.012	0.939
Optimal	314	0.224	0.108	0.959

V. Conclusion

The present paper aimed to show that there are strong simplification forces in neural networks. The simplest form is realized in terms of the prototype. However, the prototype is almost always deeply hidden behind seemingly complicated surface networks, and we need to develop a method to search for the hidden prototypes at any cost. To obtain the prototype network, we need to increase important information while simultaneously decreasing unnecessary information. Our method of potentiality, which is closely related to informational entropy, is effective at reducing potentiality or entropy, but it is very challenging to augment it in the presence of strong forces of potentiality reduction. To address this problem, we introduce two types of potentiality: synchronic and diachronic potentiality. Both of these potentialities tend to reduce the corresponding strength. However, by combining and neutralizing the synchronic potentiality with the diachronic potentiality, or by using di-synchronic potentiality, it becomes possible to augment synchronic potentiality in a diachronic way, or at least weaken its tendency to reduce. We applied this method to an artificial data set in which linear and non-linear relations were explicitly included. The results showed that di-synchronic potentiality was effective in weakening synchronic potentiality reduction during the learning process. With this method, the networks tried to produce the simplest networks, close to the supposed prototype networks, with improved generalization. This implies that we could generate networks whose generalization performance was better while simultaneously making the relations between inputs and outputs easier to understand in a linear way.

For future work, we should mention three points. First, the relationship between diachronic and synchronic potentiality has not been fully analyzed at this stage. It is necessary to examine more carefully how diachronic and synchronic potentiality should be interrelated. Second, our method has been inspired by the information-theoretic approach to neural networks, such as mutual information and the information bottleneck. We need to examine more carefully the relations between our method and these more mathematically-oriented approaches. Finally, the data set used in this paper was artificially created to examine how well our method can distinguish between linear and non-linear relations. It is naturally necessary for this method to be applied to larger-scale and more practical data sets.

References

[1] S. Watanabe, *Knowing and guessing: A quantitative study of inference and information*. New York: John Wiley and Sons Inc, 1969.

[2] P. Schneider, M. Biehl, and B. Hammer, “Adaptive relevance matrices in learning vector quantization,” *Neural computation*, vol. 21, no. 12, pp. 3532–3561, 2009.

[3] M. Biehl, B. Hammer, and T. Villmann, “Prototype-based models in machine learning,” *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 7, no. 2, pp. 92–111, 2016.

[4] F. Pourpanah *et al.*, “A review of generalized zero-shot learning methods,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 4, pp. 4051–4070, 2022.

[5] D. Alvarez Melis and T. Jaakkola, “Towards robust interpretability with self-explaining neural networks,” *Advances in Neural Information Processing Systems*, vol. 31, pp. 7775–7784, 2018.

[6] C. Yang, A. Rangarajan, and S. Ranka, “Global model interpretation via recursive partitioning,” in *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, IEEE, 2018, pp. 1563–1570.

[7] S. M. Lundberg *et al.*, “From local explanations to global understanding with explainable ai for trees,” *Nature machine intelligence*, vol. 2, no. 1, pp. 2522–5839, 2020.

[8] G. Alain, “Understanding intermediate layers using linear classifier probes,” *arXiv preprint arXiv:1610.01644*, 2016.

[9] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should i trust you?: Explaining the predictions of any classifier,” in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, ACM, 2016, pp. 1135–1144.

[10] D. Garreau and U. Luxburg, “Explaining the explainer: A first theoretical analysis of lime,” in *International Conference on Artificial Intelligence and Statistics*, PMLR, 2020, pp. 1287–1296.

[11] G. Hinton, O. Vinyals, and J. Dean, “Distilling the knowledge in a neural network,” *stat*, vol. 1050, p. 9, 2015.

[12] P. Luo, Z. Zhu, Z. Liu, X. Wang, and X. Tang, “Face model compression by distilling knowledge from neurons,” in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[13] R. Mishra, H. P. Gupta, and T. Dutta, “A survey on deep neural network compression: Challenges, overview, and solutions,” *arXiv preprint arXiv:2010.03954*, 2020.

[14] R. Linsker, “Perceptual neural organization: Some approaches based on network models and information theory,” *Annual review of Neuroscience*, vol. 13, no. 1, pp. 257–281, 1990.

[15] N. Tishby and N. Zaslavsky, “Deep learning and the information bottleneck principle,” in *2015 ieee information theory workshop (itw)*, IEEE, 2015, pp. 1–5.

[16] A. A. Alemi, I. Fischer, J. V. Dillon, and K. Murphy, “Deep variational information bottleneck,” *arXiv preprint arXiv:1612.00410*, 2016.

[17] K. Shimizu, *Multivariate analysis (in Japanese)*. Nikkan Kogyo Shinbun, 2009.

 ISBN: 978-1-68558-260-9

"Red Cars are Faster than Other Cars": The Impact of Color on Children's Estimation of Speed

Jerome Dinet
University of Lorraine, 2LPN and Chair BEHAVIOUR
Nancy, France
jerome.dinet@unuv-lorraine.fr

Mickael Smodis
MSH Lorraine
Nancy, France
mickael.smodis@univ-lorraine.fr

Robin Vivian
University of Lorraine, PErSEUs
Metz, France
robin.vivian@univ-lorraine.fr

Pierre Chevrier
Lorraine ENIM and Chair BEHAVIOUR
Metz, France
pierre.chevrier@univ-lorraine.fr

Melanie Laurent
University of Lorraine, 2LPN
Nancy, France
melanie.laurent@univ-lorraine.fr

Gaelle Nicolas
University of Lorraine and Chair BEHAVIOUR
Metz, France
gaelle.nicolas@univ-lorraine.fr

***Abstract*—Estimation of speed is a major problem for young pedestrians to make the decision about when and where it is safe to cross the road. If some authors concentrated on visual exploration of young pedestrians during crossing activity, no studies have looked at the impact of vehicle color on the estimation of speed by children. Our experiment, conducted with 67 children aged 7, is aiming to investigate the impact of four environmental factors (type of object, speed of the objects, direction of the objects, color of the objects) on the estimation of the speed of moving objects by children. Results have mainly shown that, if direction of the objects has no impact of estimation of speed, red objects (cubes or cars) are significantly perceived as the fastest. Moreover, when the objects are cars, the number of "Red" responses is significantly superior than "Blue" responses. And finally, for our participants, red objects are significantly perceived as being faster, especially if they are cars and the speed of movement is fast. Rather than thinking of pedestrian safety as a body of knowledge to absorb, it is most useful to characterize it as a series of cognitive, perceptual, and motor skills that must be mastered to navigate in the traffic environment.**

Keywords*—*Child; Pedestrian; Safety; Risk; Hazard

I. INTRODUCTION

Pedestrian trauma represents a significant proportion of all road traumas, young pedestrian being over-represented in all these road traumas. In particular, the safety of child pedestrians is of concern, given that a sizable proportion of pedestrians killed and seriously injured involve children and the special value society places on its youth [1] [2]. Because children's adaptation to an oncoming vehicle's distance and speed is a critical component of their street-crossing decision-making, several authors investigated how children experience the speed of vehicles when they must to cross a street [3][4]. If all these studies provide very interesting results demonstrating that children's behaviors are affected by many factors, no study has investigated the impact of colors for different children's age groups while some authors recognized that colors can be determinant [5].

It is the reason why this paper is aiming to investigate the impact of the red color of vehicle on decision taking and evaluation made by young pedestrians.

A. Context

Around the world, the number of pedestrians killed increase. Young pedestrians are particularly concerned by these accidents: For instance, according to the official data issued from the Traffic Safety Facts, on average, three children were killed and an estimated 502 children were injured every day in the U.S. in traffic crashes. In 2019 and 2020, there were respectively 181 and 177 children killed in pedestrian accidents. In the same way, in France, in 2022, the number of serious injuries for young pedestrians is estimated to be 5 percents higher than in 2019. Most were toddlers (between the ages of 1-3) and young children (4-7). In fact, an estimated 1 in 5 children killed in car accidents were pedestrians, i.e., just walking on the sidewalk or crossing a street whatever the country [6][7].

At ages 6-10 years, children are at highest risk of pedestrian collision [2], most likely due to the beginning of independent unsupervised travel at a time when their road strategies, skills and understanding are not yet fully developed. Whatever the country, research suggests that children between the ages of 6 to 10 are at highest risk of death and injury, with an estimated minimum four times the risk of collision compared to adult pedestrians [8]. Until the age of 6-7 years, children are under active adult supervision, i.e., parents hold their child's hand when crossing roads together. Even if every year many pedestrians are injured or killed in traffic accidents in rural parts of the country [9], pedestrian safety is being considered as a serious traffic safety problem in urban and suburban settings [10][11]. Thus, children more than adults, are at risk as pedestrians, often due to their own actions and behaviors. So, the question is: "Why do young pedestrians not adopt safety behaviors specially during street crossing?"

From a cognitive point of view, road crossing ability is a high and complex mental activity because the individual has to process dynamic and complex information from his/her surrounding environment, to make a decision, *i.e.*, where and how to cross [12][13]. Some objects automatically attract attention away from other objects in the visual field [14]. Successful crossing performance also requires reliable estimation of the pedestrian's walking speed, peak capabilities, and distance to the other side of the road or a traffic island. Integrating all these aspects is difficult for the child, especially one inexperienced in traffic, and result in a longer decision making time: In fact, a 5 year old child requires about twice as long to reach a pedestrian decision as an adult.'and This leaves even less time to execute an imperfectly planned crossing [12][13][15]. A vast amount of research suggests that children's development of cognitive skills is significantly related to increased pedestrian safety and that relevant skills improve as children get older [16][17].

Several studies showed that estimation of speed and assessment of inter-vehicular gap by young pedestrians are two major problems for young pedestrians to make the decision about when and where it is safe to cross the road [18][19][20] [21]. But, if some authors concentrated on visual exploration of young pedestrians during crossing activity [22][23], no studies have looked at the impact of vehicle color on the estimation of speed by children.

B. Are Red Car Faster than Other ?

From a statistical point of view, some data revealed that certain colored cars are more likely than others to be involved in an accident because physical explanations exist. For instance, darker colors make it harder for vehicles to be seen, especially at night, as they blend more into their surroundings, while white cars are highly visible. The belief that red cars are faster is a persistent automotive myth rooted in psychological associations rather than factual evidence. While red cars may appear more dynamic and exciting, their color does not directly influence their speed or performance.

From a psychological point of view, we can distinguish four hypotheses to explain that red cars can be perceived as faster than other cars:

- Based on the probabilistic approach, the preferred color by buyers of fastest cars is red. For instance, as Figure 1 shows, more than 30% of Lamborghini are red. So the probability to see a red Lamborghini is objectively higher to see a Lamborghini with another color. In the same way, about 45% of the cars that Ferrari sells is ordered in red. On this basis, humans generalize information learned about a subset of a category to the category itself. This cognitive bias is very well-documented specially for children [24] [25];
- Based on the social learning theory elaborated by Albert Bandura, a large number of our behaviors is learned by observing and imitating the behavior, attitudes and emotions of others. "Learning would be excessively laborious, not to say perilous, if people had only to rely on the effects of their own actions to derive information about what to do. Fortunately, most human behavior is learned through observation and 'modeling". In the case of vehicles, a series of anime and movies display red cars, these cars being very fast [26];
- Based on ethology, several studies show that red color guide attention towards important objects in nature, in particular for mammals [27]. More precisely, red has a direct and automatic influence on cognitive processes, including attention. It was established that viewing red immediately before or during a motor response increases the response's strength and velocity, most probably due to the elicitation of fear. This influence is consistent with the emotional evaluation of color as either hospitable or hostile. Fire trucks (and fire machines in general, such as extinguishers) have been painted red since the 1900s. To be distinguishable from the average road user, firehouses across the Country painted their apparatuses red to maximize visibility. In this way, the color red ensured the safety and conspicuity of firefighters and the fire trucks they were responding in. Today, to improve attention, fire trucks are additionally equipped with sirens, lights, and retroreflective markings to signal road users that they are responding to or returning from an emergency;
- Some studies have shown that red is associated with positive emotion for children [28], and Several studies have shown that red is consistently observed to be the children's favorite color, contrasting with the preference for blue in adults [29] [30] [31]. Based on a series of experiments, the research by Burkitt and colleagues has provided the best controlled evidence that the use of colors in drawings is not arbitrary, but instead reflects emotional associations and preferences for children. In their studies, children were invited to use color for drawings [32] [33] [34] [35] to characterized positive, negative or neutral emotion. Their results showed that children used their favorite colors for positive-related characters, their least favorite ones for negative-related characters, and colors with medium preference for neutral characters.

II. METHOD

Our experiment, conducted with 67 children aged 7, is aiming to investigate the impact of four environmental factors (type of object, speed of the objects, direction of the objects, color of the objects) on the estimation of the speed of moving objects by children.

A. Participants

Sixty-seven French participants were recruited to participate in this study (34 boys and 33 girls; Mean age = 7;3 years-

 ISBN: 978-1-68558-260-9

Cars© | The Duke of Hazzard© | Starsky and Hutch© | Fast and Furious© | Magnum© | Nicky Larson©

Figure 1. Some popular red cars in different movies and anime

old, SD = 0.9 months). All children come from the same elementary school located in the mid-town. All participants are French native speakers and the majority (86.4%) lives in urban area. All parents agreed to their children participate. No child has severe visual impairment and no cognitive impairment.

B. Independent and Dependent Variables

We manipulated four independent factors, all being within-subject:

- Type of objects in movement, with two modalities: Cube vs. Car;
- Speed of the objects, with three modalities: Slow, Medium, High. The corresponding selected test speeds are 30 km/h (Slow), 50 km/h (Medium), and 70 km/h (High). These speeds were chosen because, according to the ”Law on Road Traffic Safety” in France, the speed of vehicles in the school zone, in the settlement, is limited to 30 km/h, and outside the settlement to 50 km/h. The permitted speed of vehicles outside the settlement is 70 km/h;
- Direction of the objects, with two modalities: Equal, *i.e.*, going in the same direction vs. Different, *i.e.*, passing each other;
- Color, with two modalities: Blue vs. Red;

Only one dependent factor (measure) is collected: the oral response provided by the participant. Question addressed to children was: *According to you, what was the fastest object (or car)?*”. So each participant could answer ”Blue”, ”Red” or ”I don’t know”. There is no time limit to provide the answer.

We also formulate several control variables to avoid biasing the experiment, such as: Counterbalancing of videos for each participant (to avoid an order effect); Presenting the same videos; No severe cognitive impairments for our participants; No known or proven speech disorders; Correct vision declared (correction for 80% of participants); Data collection during the month of December 2024.

C. Material

We used the software Blender to create our 24 videos (Figure 2) showing objects in movement. Blender is a free, open-source software package dedicated to 3D creation. It is used in many fields, including animation, video games, virtual reality, and architectural visualization. Its popularity is based on its power, flexibility, and active community, which contributes to its continuous improvement. Blender enables all stages of 3D production: modeling, texturing, rigging, animation, rendering, simulation and even video editing. Thanks to its customizable graphical interface, users can adapt the working environment to their needs. Blender’s advanced tools also include features for sculpting, shading and compositing. The operating principle is based on a scene-based organization. A scene contains 3D objects (curves, lights, cameras, etc.) that can be manipulated in a virtual space. These objects can be modified using transformations (translation, rotation, scaling), materials and physical effects.

D. Procedure and Design

The entire experimentation could be done at a desk with the children seated. There were an experimenter (always the same) and one participant in a quiet room located in the school. The room had only desks and chairs without any distractions. The experiment has four steps:

- First, the experiment begins with a brief greeting and oral instruction. The instructions are as follows: “*You are going to see different videos of objects or cars in movement. For each video, simply tell me which car you think is the fastest*”;
- If the participant has no question, two trials are provided. Each participant is instructed to be relaxed and to gaze at the center cross during the interval screen for 5 seconds. After this 5 seconds, one of the videos was shown. The participant was instructed to answer the question orally;
- After these two trials, the experiment can begin effectively. This process was repeated with other videos again and again until the participant saw all 24 videos. Figure 3 shows the protocol used;
- Finally, each participant is thanked and a debriefing takes place to gather the child’s impressions.

Throughout the experiment, participants were allowed to quit if they felt too bad or too stressed to continue.

E. Design and Data Analysis

The overall analysis plan (AP) is given by the following relationship.:

$$AP = S \times Object_2 \times Speed_3 \times Direction_2 \times Color_2$$

F. Ethics

All legal parents of children provided the same informed consent. Moreover, the responsible of the school for children

 ISBN: 978-1-68558-260-9

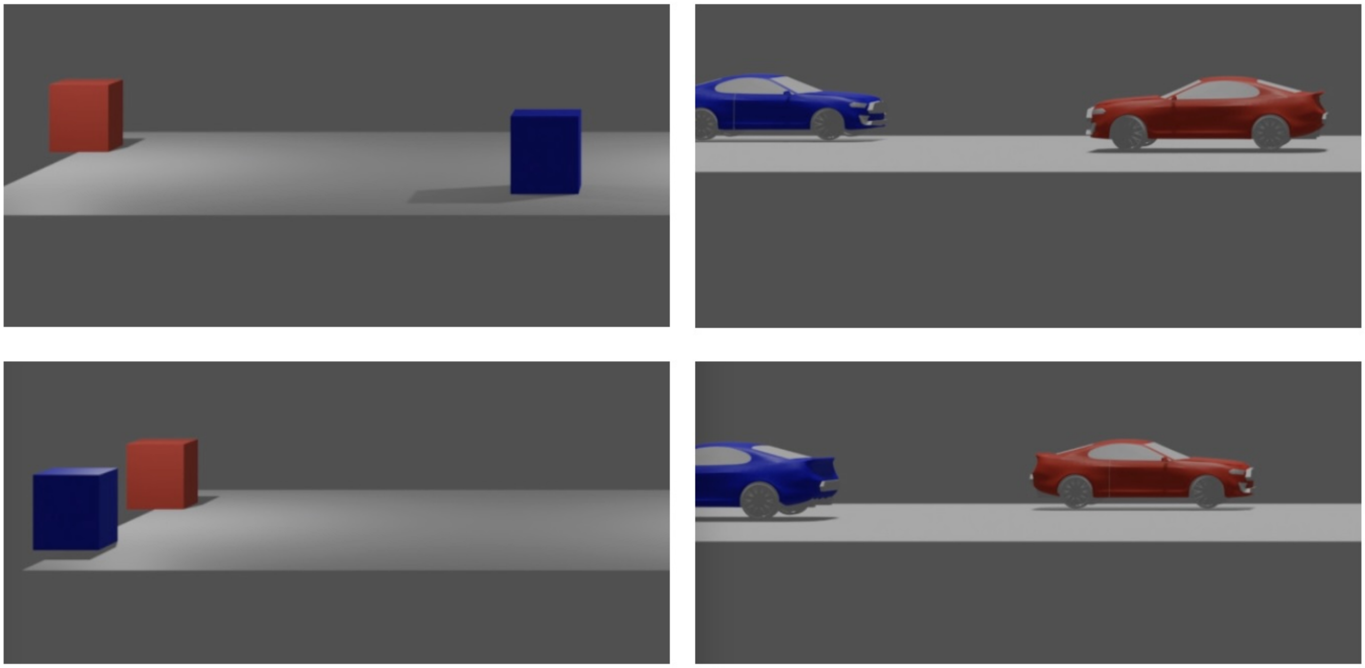


Figure 2. Screenshots of the material displayed on the screen for the experiment

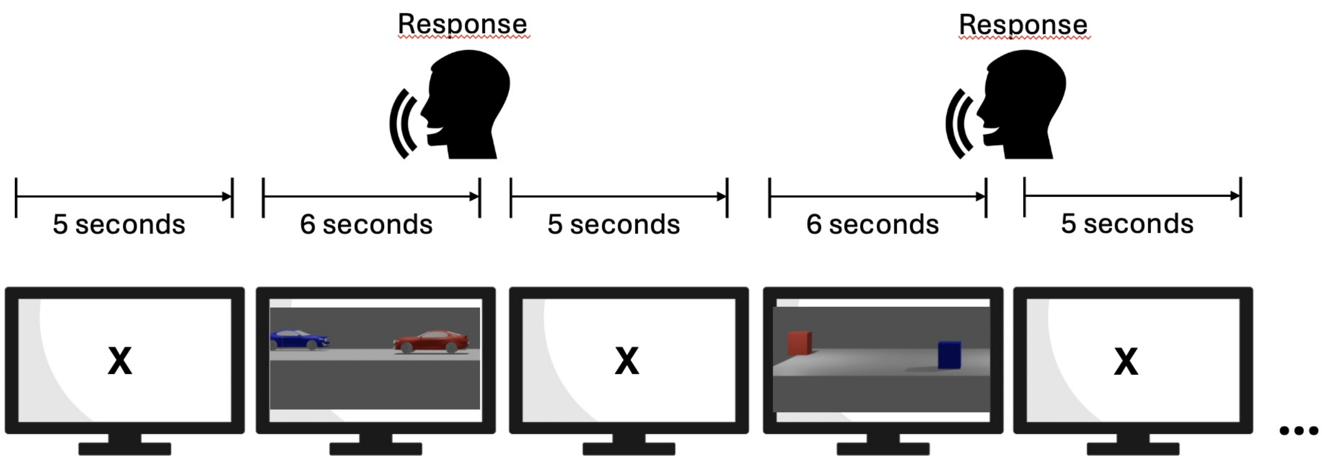


Figure 3. The procedure of the experiment

provided also her consent. All they received the same information relating to the following points:

- A statement that participation is voluntary and that refusal to participate will not result in any consequences or any loss of benefits that the person is otherwise entitled to receive;
- The precise purpose of the research;
- The procedure and material involved in the research;
- Benefits of the research to society and possibly to the individual human subject;
- Length of time the subject is expected to participate
- Researchers ensured that those participating in research will not be caused distress;
- Subjects' right to confidentiality and the right to withdraw from the study at any time without any consequences;
- After the research is over, each of them are able to discuss the procedure and the findings with the psychologist.

III. Results

As Figure 4 shows, several main effects have been obtained:

- To the question, *"According to you, what was the fastest object (or car)?*, 30% of children answered "Red", while only 12.41% answered "Blue". So red objects (cubes or cars) are significantly perceived as the fastest ($\chi^2(1, N = 66) = 17.54$, $p = .001$);

 ISBN: 978-1-68558-260-9

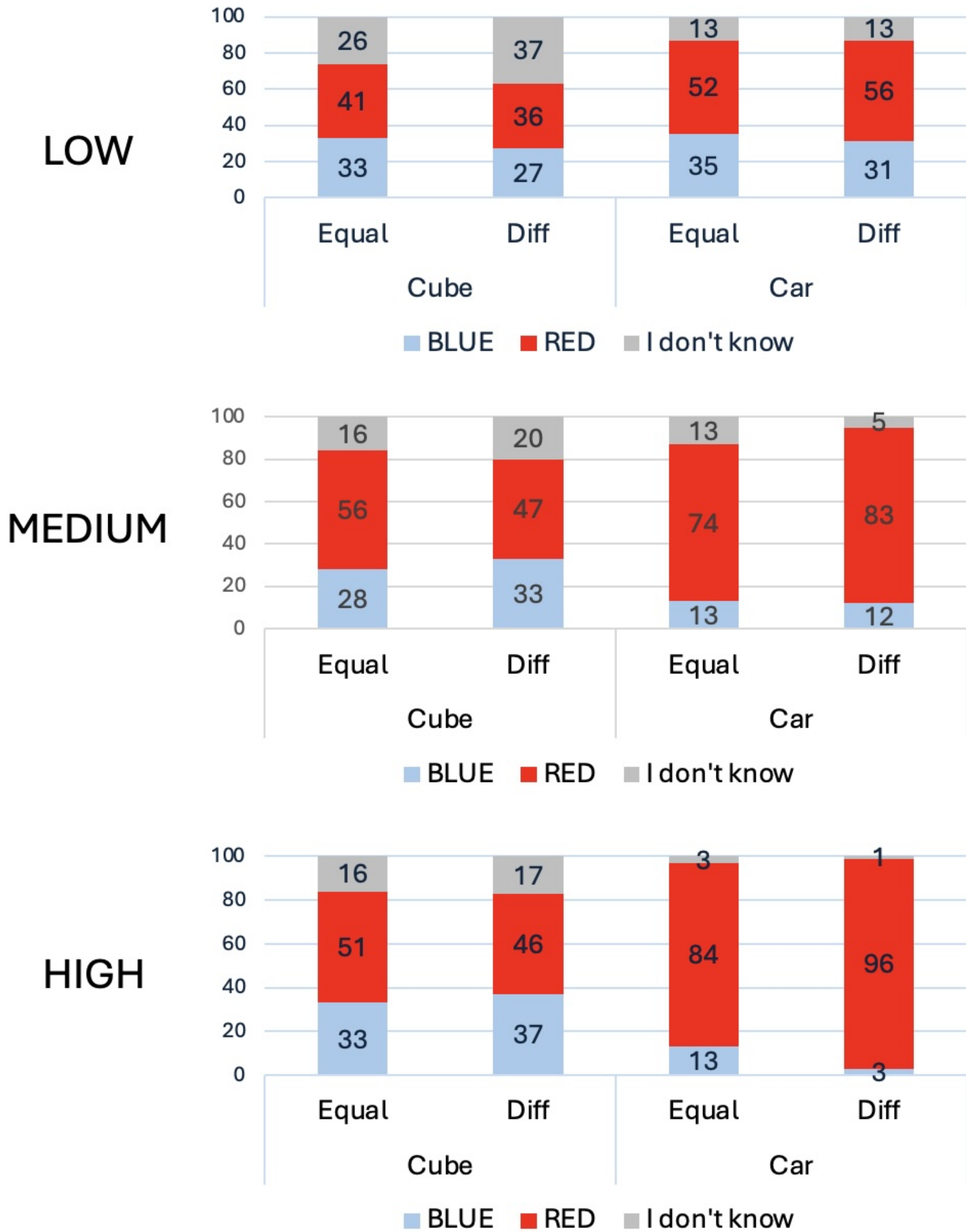


Figure 4. "According to you, what was the fastest object (or car)?" Percentages of responses given by children for the two kinds of objects (Cube vs. Car), for the two senses of direction (Equal vs. Different) and for the three speeds (Low, Medium, High)

- Type of objects in movement (Cube vs. Car) has a significant impact on responses given by children. When the objects are cubes, the number of "Red" and "Blue" responses is identical ($\chi^2(1, N = 66) = 6.37$, $p = .06$). But when the objects are cars, the number of "Red" responses is significantly superior than "Blue" responses ($\chi^2(1, N = 66) = 16.68$, $p = .004$);
- Speed of the objects has a significant impact on responses: (i) For the "Slow" speed condition (30 km/h), 46.2% of children answered "Red", while 31.5% answered "Blue" ($\chi^2(1, N = 66) = 10.23$, $p = .03$); (ii) For the "Medium" speed condition (50 km/h), 65% of children answered "Red", while 21.5% answered "Blue" ($\chi^2(1, N = 66) = 19.37$, $p = .003$); For the "High" speed condition (70 km/h), 70.25% of children answered "Red", while 20.5% answered "Blue" ($\chi^2(1, N = 66) = 26.74$, $p = .001$). In other words, the higher the speed, the more the children respond "Red";
- Direction of the objects/cars has no impact on responses given by the children, ($\chi^2(1, N = 66) = 3.21$, $p = .06$).

To summary, for our participants, red objects are significantly perceived as being faster, especially if they are cars and the speed of movement is fast.

IV. Discussion

Our experiment, conducted with 67 children aged 7, aimed to study the impact of different environmental factors (type of object, speed of the objects, direction of the objects, color of the objects) on the estimation of the speed of moving objects by these children.

If direction of the objects has no impact of estimation performed by the children, the other three factors have a significant impact: Red objects (cubes or cars) are significantly perceived as the fastest; When the objects are cars, the number of "Red" responses is significantly superior than "Blue" responses; And the higher the speed, the more the children respond "Red". Finally, for our participants, red objects are significantly perceived as being faster, especially if they are cars and the speed of movement is fast.

Several methodological limitations prevent us to generalize the results obtained. For instance, only two colors are used in our videos (red and blue) and there are no decorative elements. So, it could be interesting to reproduce this experiment by adding decorative elements and other vehicles or pedestrians because several studies have shown that the physical and social context are determinant in the behavior of young pedestrians [10][36][37][38].

Nevertheless, our experiment confirms that some environmental factors, such as the color can influence road crossing activity which is a high and complex cognitive activity because the individual has to process dynamic and complex information from his/her surrounding environment, to make a decision, *i.e.*, where and how to cross, especially when these individuals are young pedestrians. Safe pedestrians must possess and utilize advanced cognitive skills [12][13]. Crossing decisions include whether or not to enter the roadway, the place to cross, the path to take, how fast to travel, and how the driver might react. A sound decision on whether to enter the roadway should be based upon recall (experience) and monitoring of the traffic detected, including the distance, speed, and anticipated direction of vehicles and the opportunities provided by various gaps in traffic [39]. The time that has elapsed while making the decision also needs to be incorporated. Successful crossing performance also requires reliable estimation of the pedestrian's walking speed, peak capabilities, and distance to the other side of the road or a traffic island.

V. Conclusion and Future Works

All individuals, specifically the youngest, must learn how to be a pedestrian. Rather than thinking of pedestrian safety as a body of knowledge to absorb, it is most useful to characterize it as a series of cognitive, perceptual, and motor skills that must be mastered to navigate in the traffic environment. We argue that teaching children how to be safe pedestrians is crucial for their safety and well-being. Learning pedestrian skills, like using crosswalks, looking both ways before crossing the street, and being aware of speed of vehicles, should be introduced at an early age.

Acknowledgment

Thanks are due to all children, their parents and educational staff, and thanks to the experimental platform ECHO related to the MSH Lorraine (UAR3261, CNRS and Lorraine University).

References

[1] D. N. Lee, D. S. Young, and C. M. McLaughlin, "A roadside simulation of road crossing for children," Ergonomics, vol. 27, no. 12, 1984, pp. 1271–1281.

[2] J. Ma, Z. Shen, N. Wang, X. Xiao, and J. Zhang, "Developmental differences in children's adaptation to vehicle distance and speed in street-crossing decision-making," Journal of safety research, vol. 88, 2024, pp. 261–274.

[3] V. Gitelman, S. Levi, R. Carmel, A. Korchatov, and S. Hakkert, "Exploring patterns of child pedestrian behaviors at urban intersections," Accident Analysis & Prevention, vol. 122, 2019, pp. 36–47.

[4] A. Trifunović, M. Vujanić, D. Pešić, S. Čičević, and M. Čubranić-Dobrodolac, "The importance of color perception and motor skills of preschool children from the aspect of traffic safety," in IX International Conference, Road Safety in Local Community, 2014, pp. 473–478.

[5] A. Trifunović, D. Pešić, M. Petrović, and S. Čičević, "How children experience the speed of a vehicle?"

[6] E. K. Adanu, R. Dzinyela, and W. Agyemang, "A comprehensive study of child pedestrian crash outcomes in ghana," Accident Analysis & Prevention, vol. 189, 2023, p. 107146.

[7] K. Koekemoer, M. Van Gesselleen, A. Van Niekerk, R. Govender, and A. B. Van As, "Child pedestrian safety knowledge, behaviour and road injury in cape town, south africa," Accident Analysis & Prevention, vol. 99, 2017, pp. 202–209.

[8] M. Struik et al., "Pedestrian accident project report no. 4: Literature review of factors contributing to pedestrian accidents," Tech. Rep., 1988.

[9] J. N. Ivan, P. E. Garder, and S. S. Zajac, Finding strategies to improve pedestrian safety in rural areas. United States. Dept. of Transportation, 2001.

[10] C. DiMaggio and M. Durkin, "Child pedestrian injury in an urban setting descriptive epidemiology," Academic emergency medicine, vol. 9, no. 1, 2002, pp. 54–62.

[11] W. W. Hunter, J. C. Stutts, W. E. Pein, C. L. Cox et al., "Pedestrian and bicycle crash types of the early 1990's," Turner-Fairbank Highway Research Center, Tech. Rep., 1996.

[12] J. D. Demetre, “Applying developmental psychology to children’s road safety: Problems and prospects,” Journal of applied developmental psychology, vol. 18, no. 2, 1997, pp. 263–270.
[13] J. D. Demetre et al., “Errors in young children’s decisions about traffic gaps: Experiments with roadside simulations,” British Journal of psychology, vol. 83, no. 2, 1992, pp. 189–202.
[14] C. R. Guibal and B. Dresp, “Interaction of color and geometric cues in depth perception: When does “red” mean “near”?” Psychological research, vol. 69, no. 1, 2004, pp. 30–40.
[15] J. L. Schofer et al., “Child pedestrian injury taxonomy based on visibility and action,” Accident Analysis & Prevention, vol. 27, no. 3, 1995, pp. 317–333.
[16] T. Pitcairn and T. Edlmann, “Individual differences in road crossing ability in young children and adults,” British Journal of Psychology, vol. 91, no. 3, 2000, pp. 391–410.
[17] D. C. Schwebel, A. L. Davis, and E. E. O’Neal, “Child pedestrian injury: A review of behavioral risks and preventive strategies,” American journal of lifestyle medicine, vol. 6, no. 4, 2012, pp. 292–302.
[18] M. Congiu et al., “Child pedestrians: Factors associated with ability to cross roads safely and development of a training package,” Victoria: Monash University Accident Research Centre (MUARC), 2008.
[19] J. A. Oxley, E. Ihsen, B. N. Fildes, J. L. Charlton, and R. H. Day, “Crossing roads safely: an experimental study of age differences in gap selection by pedestrians,” Accident Analysis & Prevention, vol. 37, no. 5, 2005, pp. 962–971.
[20] J. Oxley, B. Fildes, E. Ihsen, J. Charlton, and R. Day, “Simulation of the road crossing task for older and younger adult pedestrians: a validation study,” in ROAD SAFETY RESEARCH AND ENFORCEMENT CONFERENCE, 1997, HOBART, TASMANIA, AUSTRALIA, 1997.
[21] G. Simpson, L. Johnston, and M. Richardson, “An investigation of road crossing in a virtual environment,” Accident Analysis & Prevention, vol. 35, no. 5, 2003, pp. 787–796.
[22] D. Whitebread and K. Neilson, “The contribution of visual search strategies to the development of pedestrian skills by 4-11 year-old children,” British journal of educational psychology, vol. 70, no. 4, 2000, pp. 539–557.
[23] H. Tapiro, A. Meir, Y. Parmet, and T. Oron-Gilad, “Visual search strategies of child-pedestrians in road crossing tasks,” Proceedings of the Human Factors and Ergonomics Society Europe, 2014, pp. 119–130.
[24] S. L. Sutherland, A. Cimpian, S.-J. Leslie, and S. A. Gelman, “Memory errors reveal a bias to spontaneously generalize to categories,” Cognitive science, vol. 39, no. 5, 2015, pp. 1021–1046.
[25] T. B. Ward, A. H. Becker, S. D. Hass, and E. Vela, “Attribute availability and the shape bias in children’s category generalization,” Cognitive Development, vol. 6, no. 2, 1991, pp. 143–167.
[26] M. Rondier, “A. bandura. auto-efficacité. le sentiment d’efficacité personnelle. paris: Éditions de boeck université, 2003,” L’orientation scolaire et professionnelle, no. 33/3, 2004, pp. 475–476.
[27] M. Kuniecki, J. Pilarczyk, and S. Wichary, “The color red attracts attention in an emotional context. an erp study,” Frontiers in human neuroscience, vol. 9, 2015, p. 212.
[28] C. J. Boyatzis and R. Varghese, “Children’s emotional associations with colors,” The Journal of genetic psychology, vol. 155, no. 1, 1994, pp. 77–85.
[29] P. Valdez and A. Mehrabian, “Effects of color on emotions.” Journal of experimental psychology: General, vol. 123, no. 4, 1994, p. 394.
[30] S. Gil and L. Le Bigot, “color and emotion: children also associate red with negative valence,” Developmental science, vol. 19, no. 6, 2016, pp. 1087–1094.
[31] M. R. Zentner, “Preferences for colors and color–emotion combinations in early childhood,” Developmental Science, vol. 4, no. 4, 2001, pp. 389–398.
[32] E. Burkitt, K. Tala, and J. Low, “Finnish and english children’s color use to depict affectively characterized figures,” International Journal of Behavioral Development, vol. 31, no. 1, 2007, pp. 59–64.
[33] E. Burkitt, M. Barrett, and A. Davis, “Children’s color choices for completing drawings of affectively characterised topics,” Journal of child psychology and psychiatry, vol. 44, no. 3, 2003, pp. 445–455.
[34] E. Burkitt and L. Sheppard, “Children’s color use to portray themselves and others with happy, sad and mixed emotion,” Educational Psychology, vol. 34, no. 2, 2014, pp. 231–251.
[35] E. Burkitt and D. Watling, “How do children who understand mixed emotion represent them in freehand drawings of themselves and others?” Educational Psychology, vol. 36, no. 5, 2016, pp. 935–955.
[36] T. Foulsham, E. Walker, and A. Kingstone, “The where, what and when of gaze allocation in the lab and the natural environment,” Vision research, vol. 51, no. 17, 2011, pp. 1920–1931.
[37] E. Crawford, J. Gross, T. Patterson, and H. Hayne, “Does children’s color use reflect the emotional content of their drawings?” Infant and Child Development, vol. 21, no. 2, 2012, pp. 198–215.
[38] D. Farrelly, R. Slater, H. R. Elliott, H. R. Walden, and M. A. Wetherell, “Competitors who choose to be red have higher testosterone levels,” Psychological science, vol. 24, no. 10, 2013, pp. 2122–2124.
[39] R. A. Schieber and N. Thompson, “Developmental risk factors for childhood pedestrian injuries.” Injury Prevention, vol. 2, no. 3, 1996, p. 228.

Effects of Experience of Listening to Short Sentences Containing ANEWs on Memory: An Analysis Based on Pupillary Responses during Listening and Visual Behavior during Impression Evaluation

Katsuko T. Nakahira
Nagaoka University of Technology
Nagaoka, Niigata, Japan
Email: katsuko@vos.nagaokaut.ac.jp

Sho Hasegawa
Nagaoka University of Technology
Nagaoka, Niigata, Japan
Email: s211070@stn.nagaokaut.ac.jp

Shunsuke Moriya
Nagaoka University of Technology
Nagaoka, Niigata, Japan
Email: s213345@stn.nagaokaut.ac.jp

Muneo Kitajima
Nagaoka University of Technology
Nagaoka, Niigata, Japan
Email: mkitajima@kjs.nagaokaut.ac.jp

***Abstract*— The present study focuses on the memory of verbal information by short sentences provided auditorily and examines the relationship between the attributes of emotion-inducing words, i.e., Affective Norms for English Words (ANEW), in the verbal information and the memory of the information contained in the short sentences. We have developed a cognitive model of auditory information comprehension, in which auditory information is sequentially input from the outside, stored in working memory, cognitively understood, and the memory network is updated to reflect the results of these processes. According to this model, the content retained in working memory during listening should influence subsequent recall, as externally provided auditory information is transient. Furthermore, previous studies have shown that the presence of ANEWs also affects memory performance. In the current study, 32 short sentences (each read aloud within 15-20 seconds) were created by embedding two ANEWs—one evoking a positive emotion and the other a negative emotion—within four different patterns. The interval between the presentation of the first and second ANEW varied: either 1 second or 7 seconds. Seventeen subjects listened to these short sentences, recalled their content, and we recorded the number of items recalled. The eye-tracker was used to measure pupillary responses during listening and visual behavior during the selection of evaluation items related to the impression of the listening experience. We analyzed the relationship between total pupil diameter change, number of fixations, impression evaluation time required, impression evaluation results, interval between ANEWs, positive/negative ANEW combination patterns, and recall performance. The following suggestions for the mechanism leading to high recall performance were obtained: 1) when the way ANEW pairs appear in a short sentence increases the load on their cognitive processing; 2) when the cognitive processing load is not high but the impression of short sentences are strong. Finally, we discussed ways to apply the findings from this study to the design of auditory information that facilitates memory.**

Keywords*— *Affective Norms for English Words; Cognitive Process; Two Minds; Pupillary Response; Eye Movement; Recall;

I. Introduction

In our daily lives, we receive physical and chemical stimuli from the outside world through our sensory organs and process them cognitively by combining them with various information stored in the brain. To understand these processes, information obtained from perceptual processes is very useful. In particular, suggestions on the cognitive processes provided by visual information, i.e., eye movement and pupil measurements, are important and have been reviewed by many researchers.

For example, Mahanama et al. [1] summarize the methods for measuring eye gaze and pupil response. Vilotijević [2], for example, suggests that the intensity of pupillary response is influenced not only by the physical characteristics of light but also by cognitive factors such as gaze.

Studies on cognitive load, eye movement, and pupillary response, for example, include the following. Skaramagkas et al. [3] focus on visual attention, emotional arousal and cognitive workload, presenting a review of metrics related to gaze and pupil tracking (gaze, fixations, saccades, blinks, pupil size variation, etc.) that are used to detect emotional and cognitive processes. Other studies include one that examined whether eye movements respond to the viewer's cognitive load, and another that examined pupil dilation as a measure of lexical retrieval.

Research on pupillary response and emotion (valence-arousal pair) is also ongoing. In the past, pupil size variation during and after auditory emotional stimulation was examined by [4]. They suggested that pupil dilation is greater for emotionally negative/positive stimuli than for neutral stimuli. Henderson et al. [5] investigated pupil response during emotional imagery, and found that pupil dilation during emotional (pleasant or unpleasant) imagery was significantly greater than pupil dilation during neutral imagery. They suggested that pupil diameter was increased during emotional (pleasant or unpleasant) imagery compared to neutral imagery. Oliva et al. [6] investigated the relationship between pupil response and the process of emotion recognition, and showed that during emotion recognition, the time course of pupil response is driven by the decision-making process. Tarnowski et al. [7] investigated emotion recognition using eye tracking. They calculated 18 features related to eye movements (fixations and saccades) and pupil diameter, and found up to 80% classification accuracy for high arousal and low valence, low arousal and middle valence, high arousal and high valence, and low arousal and low valence, and high arousal and high valence.

Thus, eye movement and pupil response have the potential to detect the occurrence of emotion in response to emotional stimuli input to humans. Based on the above, we have been continuously conducting basic research on designing multi-modal content that can be easily recalled by real-time detection of such emotional responses [8][9][10][11][12]. So far, we have suggested the following for content with visual-auditory multimodal information: 1) It is possible to optimize the load of perception information processing by adjusting the interval between perception information. 2) The degree of emotion generated by Affective Norms for English Words (ANEW) may suppress the characteristics of the total change in pupillary response. 3) We constructed a cognitive model of the process of sequential input of externally provided auditory information, its retention in working memory, its recognition, and its updating in the memory network.

However, previous experiments have not yet investigated the relationship between the impression evaluation of the presented stimulus, the total change in pupillary response, and the recall rate after listening to the presented stimulus. For this reason, this paper investigates the relationship among eye movement data, total change in pupillary response, and recall rate at the time of selecting rating items regarding the impression of listening experience, based on pupillary response data during short sentence listening and experiences with multiple positive and negative two-level valences. We investigate the relationship between the total change in pupillary response and recall rate.

In Section II, we describe the impression evaluation process and possible measurements based on the cognitive model presented in [12]. In Section III, we describe the experiment and its analysis results. In Section IV, we discuss the relationship between pupillary response to ANEW and recall.

II. Valence Evaluation Process and Possible Biometric Mesurments

In this section, based on the cognitive model of listening to auditory information proposed by Nakahira et al. [12], possible biometric measurements are selected for impression evaluation after listening to auditory information.

Figure 1 shows the main flow. The response model for human emotion shown in Figure 1 is the Construction-Integration Model (CI-model) proposed by Kintsch [13][14][15]. The CI-model is a theory of discourse comprehension consisting of a construction step and an integration step, in which symbolic features are integrated using the connectionist technique. The cognitive model was constructed as follows, incorporating this theory. The presented auditory information is matched with long-term memory for each specific packet of sound waves, which are basically groups of phonemes that correspond to specific concepts. The matching results are returned together with the valence and other information associated with the concept. The returned matching results stay in working memory as a group for a certain period of time. The more the concepts overlap in working memory, the stronger the connection between the concepts becomes. A series of concepts are returned to long-term memory as related, but in this case, various node relationships, such as between concepts and between concepts and valences, are strengthened.

The possibility of measuring human data by applying Figure 1 is examined based on the overall picture of the pupil response shown in Nakahira et al. [12]. Since the presented stimuli assumed in this paper are sentences provided by auditory stimuli, the nature and arrangement of ANEWs in the narration are considered to influence the listener. Therefore, the control parameters are valence and arousal, which indicate the nature of ANEWs, and the interval when there are multiple ANEWs. In this paper, we deal specifically with valence among the properties of ANEWs. In this paper, two ANEWs are prepared. If positive valence is denoted as V_{++} and negative valence as V_{--}, then a pair of ANEWs can be denoted as $(V_{++},\ V_{--})$. The interval between ANEWs is denoted by dT.

The effect of these ANEWs appears in the pupil diameter of the listener, and the timing of the effect is considered to occur during and after listening to the text, respectively. In this study, we measure the total change in pupillary response, r_{all}, as proposed in Nakahira et al. [12]. The r_{all} measured each time an ANEW appears is considered a response to local emotion. In this study, since we deal with two ANEWs, we denote them as r^1_{all} and r^2_{all}, respectively. In addition, at the end of the sentence, the impression, or emotion, of the whole sentence heard should be generated, and the resulting total change in pupillary response is expressed as a r^e_{all}.

Next, the following parameters are considered measurable during impression evaluation.

- **Impression evaluation speed**:
  The speed of impression evaluation is affected by the clarity of the impression of the listening sentence. The time taken by subject p to evaluate the impression of the presented stimulus i is measured and is denoted as t^i_p.
- **Visual behavior during impression evaluation**:
  The choices when the subject responds to the impression evaluations are given as visual information. We consider the viewpoint movement during the viewing of options as an element related to the decision-making process of selecting options. The number of stopping points $n^i_{p,f}$ of p for i can be regarded as the number of times the subject searches for alternatives. The jth stop time $t^{i,j}_{f,p}$ during choice browsing indicates the approximate degree to which the user has thought about the choices.

In addition to the above measurements, the following analysis is possible.

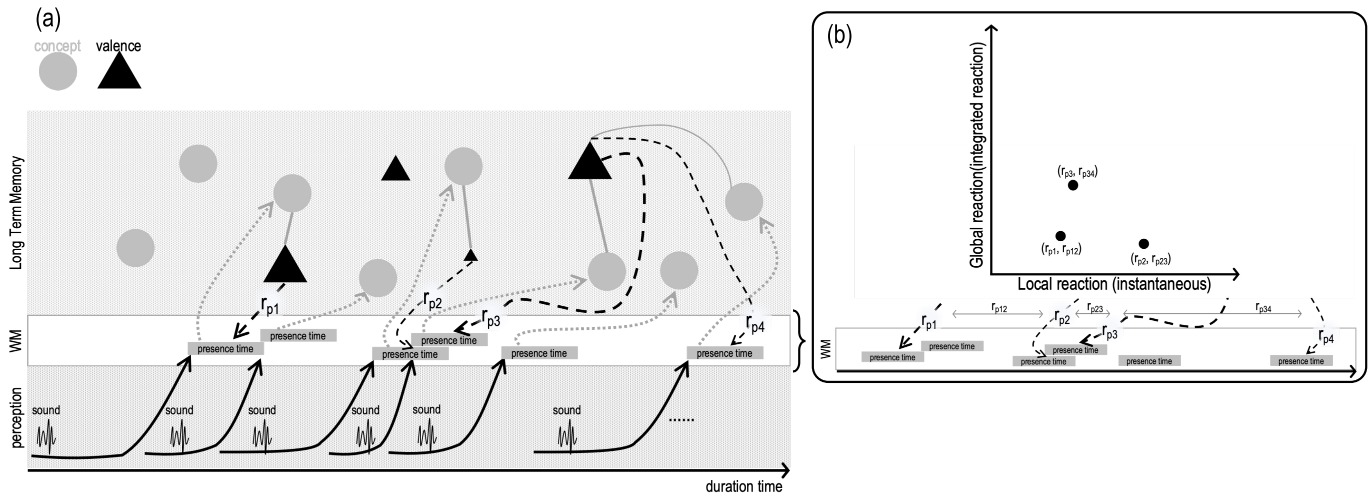


Figure 1. Cognitive model of this paper based on CI model. (a) Input - Cognitive process. (b) Working memory processing - output process. Reposting of the figure in [12].

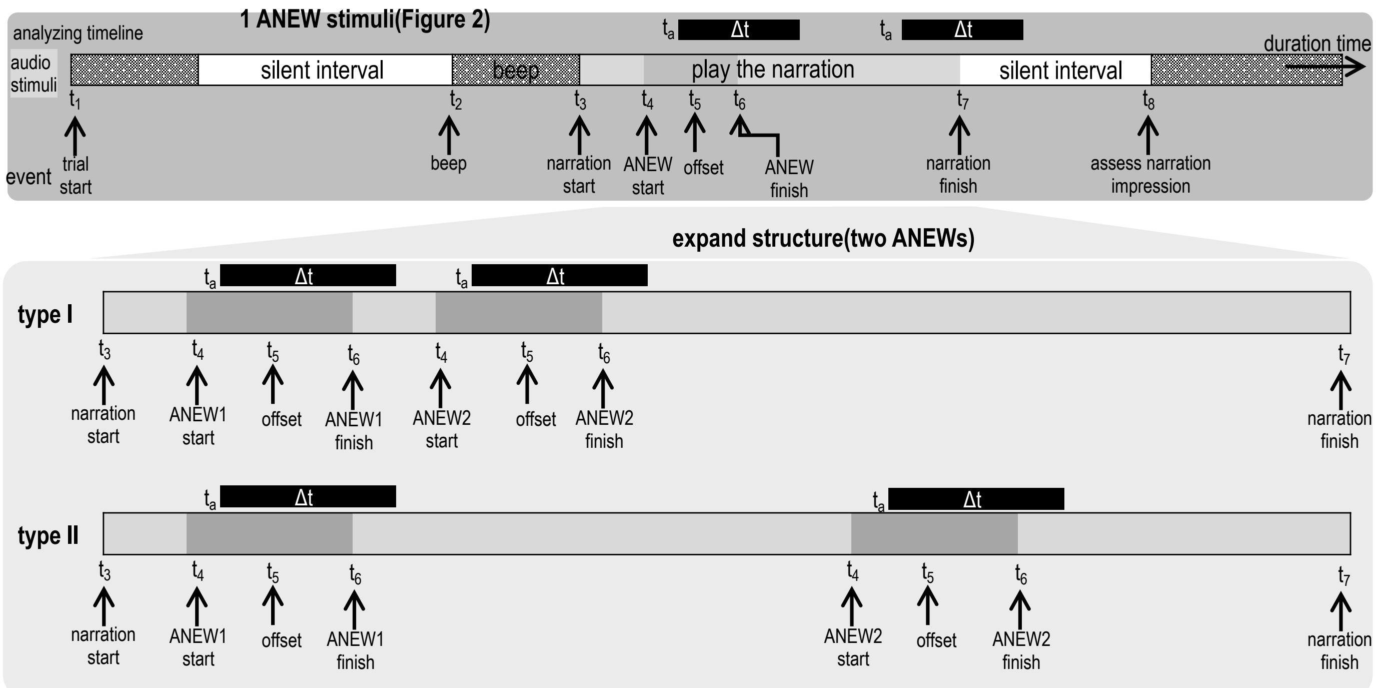


Figure 2. Timeline of presented stimuli including multiple ANEWs. Reposting of the figure in [12].

A. *Total Fixation Time during Valence Evaluation*

In the impression evaluation, it is difficult to assess the degree of concentration on the choices, i.e., the degree of thinking, from t_p^i alone. Therefore, to confirm the degree to which the subject concentrated on checking the options during t_p^i, the total pause time during the impression evaluation, $t_{f,p}^i$, is determined. This is given by the following equation:

$$t_{f,p}^i = \sum_j^{n_f^i} t_{f,p}^{i,j}$$

B. $T_{c,p}^i : t_p^i - t_{f,p}^i$ *ratio*

t_p^i and $t_{f,p}^i$ are measured in actual time required. However, when comparing only the actual time, it is difficult to distinguish between cases where the impression evaluation takes a relatively long time and those where it takes time only under certain conditions. To solve this problem, the ratio of t_p^i to $t_{f,p}^i$, $T_{c,p}^i$, is calculated as one of the indices and used as one of the characteristic quantities of impression evaluation. $T_{c,p}^i$ is calculated using the following equation:

TABLE I. PARAMETERS PREPARED IN THIS PAPER.

valence pair	dT	
	1 sec	7 sec
$(V_{++},\ V_{++})$	4	4
$(V_{--},\ V_{--})$	4	4
$(V_{++},\ V_{--})$	4	4
$(V_{--},\ V_{++})$	4	4

$$T^i_{c,p} = \frac{t^i_{f,p}}{t^i_p}$$

III. EXPERIMENT AND ANALYSIS

A. Preliminary Experiment

In order to evaluate the validity of the content reviewed in Section II, the same experiment was conducted on 17 participants using the experimental technique in multiple ANEWs indicated in Nakahira et al. [12]. The tasks of the participants were as follows. The participants were presented with 32 sentences with various valence pairs and dTs, eight sentences at a time, at random. The participants listen to the presented narration, and after each sentence, they evaluate their impression of the narration in seven levels from positive to negative and in eight levels from "no impression" to "impression unknown". The results of the impression evaluation are recorded by selecting an option displayed on the terminal and inputting the result with the mouse operation. After that, a recall test is given for each of the eight sentences, and the participant verbally states the contents of the recall. The experiment was conducted in a quiet room under constant illumination to prevent disturbance by other stimuli.

The parameters used in the experiment were as follows: first, the combinations of the ANEW and dT values are shown in the table, and the number of presented stimuli is 32 sentences. The values of valence for V_{++} ranged from 7.70 to 9.00. The values of valence for V_{--} were selected from 1.00 to 1.99. The values of valence for V_{++} were from 7.70 to 9.00, and the values of valence for V– were from 1.00 to 1.99. The number of syllables per presented stimulus was around 65. To examine the relationship between memory and the number of syllables, a recall test was conducted on the content of each sentence after every eight sentences were listened to, and the degree of recall was checked.

The following is an analysis of the parameters set in Section 2 based on the data obtained in the experiment.

B. Basic Statistics for Metrics

First, the overall picture of the data obtained in this study is shown in table II. In the table, σ indicates the standard deviation, $Q_{1/4}$ indicates the first quartile, $Q_{2/4}$ indicates the median, and $Q_{3/4}$ indicates the third quartile. Table II shows the statistics of the available data without any particular classification. The following can be read from the table.

The difference between t^i_p and $t^i_{f,p}$ in the time taken to evaluate impressions is close to one second. From this fact, it is expected that some trends in the time-related quantities can be seen by classifying the response results. For r_{all}, it is difficult to find a large difference between r^1_{all} and r^2_{all}. On the other hand, a certain difference is observed between r^e_{all} and $r^{1/2}_{all}$. This is expected to provide some insight into the amount of $r^{1/2}_{all}$ starting from r^e_{all}. We made relation analysis based on metrics $t^i_p,\ t^i_{f,p},\ T^i_{c,p},\ n^i_{p,f}$ and belows:

- valence pair and dT
- r_{all}
- results of recall test
- results of S_p

TABLE II. BASIC STATISTICS FOR THE ENTIRE DATA OBTAINED IN THE PRELIMINARY EXPERIMENT.

parameter	average	σ	$Q_{1/4}$	$Q_{2/4}$	$Q_{3/4}$
t^i_p[ms]	3958.69	2188.11	2529.50	3411.50	4666.50
$t^i_{f,p}$[ms]	2933.94	2590.75	1138.39	2220.33	4140.00
$T^i_{c,p}$	0.67	0.32	0.4	0.7	1
$n^i_{p,f}$	7.54	5.61	4	7	10
r^1_{all}[mm]	2.68	2.53	0.83	1.92	3.66
r^2_{all}[mm]	2.75	2.47	0.89	2.02	3.82
r^e_{all}[mm]	2.43	2.48	0.73	1.53	3.33

C. Relation between $t^i_p,\ t^i_{f,p},\ T^i_{c,p}, n^i_{p,f}$ and valence pair or dT

Figure 3 (a) $\sim$ (d) shows boxplots for $T^i_{c,p}$ and $n^i_{p,f}$, the valence pair, and dT. (a) in the figure shows boxplots for the valence pair and $T^i_{c,p}$, (b) in the figure shows boxplots for the valence pair and $n^i_{p,f}$, (c) in the figure shows boxplots for dT and $T^i_{c,p}$, and (d) in the figure shows boxplots for dT and $n^i_{p,f}$. Here, the analysis for the value pair is considered as follows: Normally, the valence pair should be considered only in terms of the value of the valence for ANEW. However, in the experiment in this paper, there is a possibility that the valence of one pair strengthens or weakens the valence of the other pair as a result of their interaction. For this reason, we classified the valence pairs by including dT.

Based on the figure, we performed a one-way ANOVA for the valence pairs $t^i_p,\ t^i_{f,p}, T^i_{c,p}$, and $n^i_{p,f}$. The results showed a significant difference in $T^i_{c,p}$ $(F(7,536) = 3.018, p = 0.004)$. Multiple Comparison Procedure using Tukey's method revealed significant differences in the following combinations:

- $(V_{++},\ V_{++},\ dT = 1) - (V_{--},\ V_{--},\ dT = 1)$,
- $(V_{++},\ V_{++},\ dT = 1) - (V_{--},\ V_{--},\ dT = 7)$,
- $(V_{++},\ V_{++},\ dT = 1) - (V_{--},\ V_{++},\ dT = 7)$

D. Relation between $t^i_p,\ t^i_{f,p},\ T^i_{c,p}, n^i_{p,f}$ and r_{all}

Next, a one-way ANOVA was performed for each of $r^{1/2/e}_{all}$ with the valence pair and dT. However, no significant results were obtained in either case. Therefore, a test of no correlation was performed to confirm the existence of correlation between $r^{1/2/e}_{all}$ and $t^i_p, t^i_{f,p}, T^i_{c,p}$, and $n^i_{p,f}$. As a result, the following correlations were confirmed:

- $(V_{++},\ V_{++},\ dT = 1)$
  correlation coefficient $\rho = -0.24$ $(p = 0.05)$ between $r^1_{all} - r^e_{all}$ and $T^i_{c,p}$

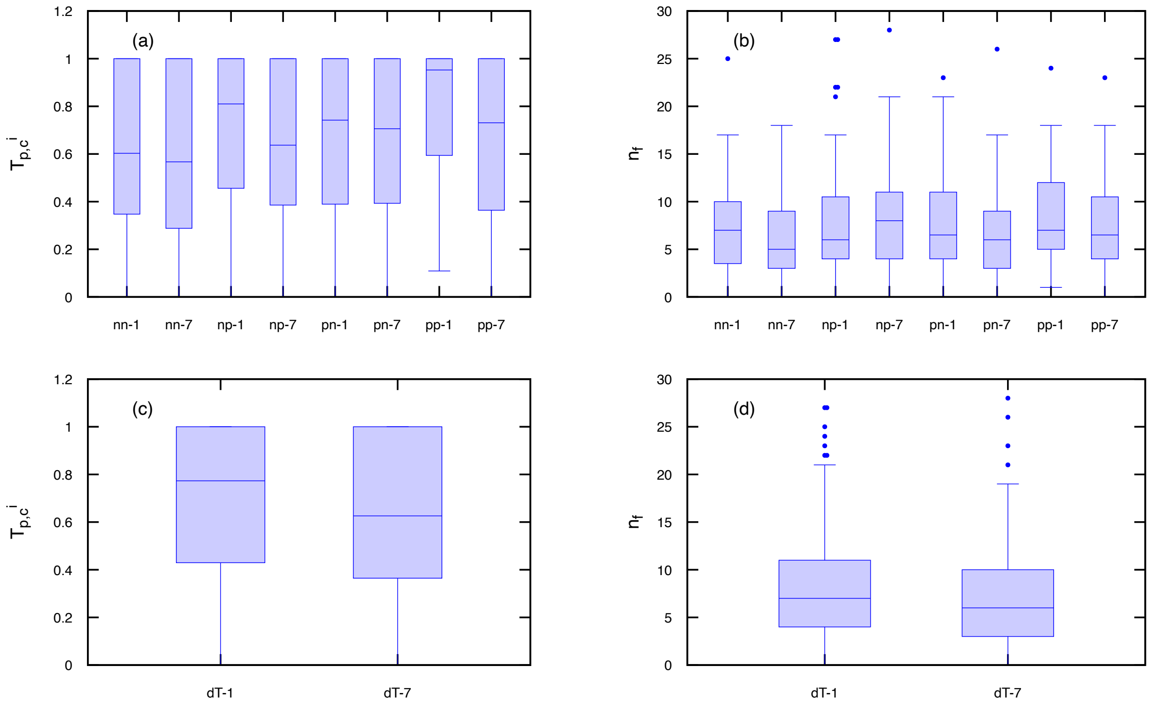


Figure 3. Boxplots for $T^i_{c,p}$ and n_f, with $T^i_{c,p}$ on the left and nf on the right. Positive valence is represented by **p**, negative valence by **n**, and the interval between the combination and ANEW is represented by a number.

TABLE III. NUMBER OF CORRECT ANSWERS FOR RECALL WITH ANEW1, ANEW2, AND dT COMBINATIONS.

$dT = 1$			
		valence type of ANEW1	
		V_{++}	V_{--}
valence type of ANEW2	V_{++}	**34**	23
	V_{--}	27	21
$dT = 7$			
valence type of ANEW2	V_{++}	*17*	25
	V_{--}	21	**28**

- $(V_{++},\ V_{++},\ dT = 7)$
  correlation coefficient $\rho = -0.237$ $(p = 0.05)$ between $r^1_{all} - r^2_{all}$ and $T^i_{c,p}$

However, all of them are weak correlation coefficients.

E. Relation between t^i_p, $t^i_{f,p}$, $T^i_{c,p}, n^i_{p,f}$ and result of the recall test

All the results have been analyzed without considering the participants' reactions. Next, we conduct an analysis that takes the participant's responses into account. First, we analyze the relationship between the results of the recall test and the various quantities. The results of the recall test were recorded by removing the recency effect and the primacy effect. In this paper, a recall test was conducted for each of the eight presented stimuli. Therefore, the results of the first and eighth sentences were excluded from the number of recalled sentences, and the sentences that were able to reproduce the meaning of the sentence were judged to be "recalled". The results are shown in Table III.

A one-way ANOVA was performed on the measured quantities according to this classification, but no significant differences were found. The results of the test of no correlation on the values of the quantities with respect to the value pairs showed the correlations shown in Table IV.

F. Relation between t^i_p, $t^i_{f,p}$, $T^i_{c,p}$, $n^i_{p,f}$, $r^{1/2/e}_{all}$ and Result of Impression Evaluation

Finally, we analyze the relationship between the classification of the subjects' impression ratings and the various measurements. In Nakahira et. al. [12], the results of the participants' impression ratings of S_p were classified according to the distance from the ANEW valence contained in the presented stimulus as follows:

$$\begin{array}{rl}
\text{VES}: & |V_1 - S_p| < 1,\ |V_2 - S_p| < 1 \\
\text{V1ES}: & |V_1 - S_p| < 1,\ |V_2 - S_p| \geq 2 \\
\text{V1LS}: & 1 \leq |V_1 - S_p| < 2,\ |V_2 - S_p| \geq 2 \\
\text{V2ES}: & |V_1 - S_p| \geq 2,\ |V_2 - S_p| < 1 \\
\text{V2LS}: & 1 \leq |V_2 - S_p| < 2,\ |V_1 - S_p| \geq 2 \\
\text{N/A}: & |V_1 - S_p| \geq 2,\ |V_2 - S_p| \geq 2
\end{array}$$

This paper also classifies S_p accordingly. An ANOVA was conducted on the various measures with S_p as the explanatory variable, but no significant differences were obtained. Strictly speaking, a one-way ANOVA showed significant differences when S_p was classified as V1ES/V1LS, but the number of samples in these classifications was less than 20, which was not precise enough to perform ANOVA. For this reason, this

TABLE IV. RESULTS OF THE TEST OF NO CORRELATION BETWEEN MEASUREMENTS FOR EACH CORRECT/INCORRECT ANSWER IN THE RECALL.

memory availability	valence pair	measurements			ρ	$p-value$
entire set	$(V_{++},\ V_{--}, dT=1)$	r^2_{all}	and	n_f	0.313	0.009
	$(V_{++},\ V_{++},\ dT=1)$	$r^1_{all}-r^e_{all}$	and	$T^i_{c,p}$	-0.240	0.049
	$(V_{++},\ V_{++},\ dT=7)$	r^1_{all}	and	n_f	0.303	0.012
		r^2_{all}	and	n_f	0.258	0.034
recalled	$(V_{++},\ V_{--}, dT=1)$	r^2_{all}	and	$t^i_{f,p}$	0.477	0.026
		r^2_{all}	and	n_f	0.534	0.011
	$(V_{--},\ V_{++},\ dT=7)$	$r^1_{all}-r^e_{all}$	and	$t^i_{f,p}$	0.564	0.008
		$r^1_{all}-r^e_{all}$	and	$T^i_{c,p}$	0.543	0.011
not recalled	$(V_{--},\ V_{++},\ dT=1)$	r^2_{all}	and	n_f	0.302	0.040
	$(V_{++},\ V_{++},\ dT=1)$	$r^1_{all}-r^e_{all}$	and	$T^i_{c,p}$	-0.351	0.024
	$(V_{++},\ V_{++},\ dT=7)$	r^1_{all}	and	n_f	0.323	0.017
		r^2_{all}	and	n_f	0.347	0.288
		$r^1_{all}-r^e_{all}$	and	n_f	0.275	0.044
		$r^2_{all}-r^e_{all}$	and	n_f	0.310	0.023

TABLE V. RESULTS OF TEST OF NO CORRELATION BETWEEN MEASUREMENTS FOR EACH TYPE OF S_p.

assess type of S_p	measurments			ρ	$p-value$
VES	r^1_{all}	and	n_f	0.21	0.011
V2ES	$r^1_{all}-r^e_{all}$	and	t^i_p	0.258	0.005
	$r^1_{all}-r^e_{all}$	and	$t^i_{f,p}$	0.298	0.001
	$r^1_{all}-r^e_{all}$	and	n_f	0.228	0.014
	r^1_{all}	and	$t^i_{f,p}$	0.236	0.010
	r^1_{all}	and	n_f	0.222	0.016
V2LS	$r^1_{all}-r^2_{all}$	and	n_f	-0.219	0.038
	$r^1_{all}-r^e_{all}$	and	t^i_p	-0.298	0.004
	$r^1_{all}-r^e_{all}$	and	$t^i_{f,p}$	-0.321	0.002
	$r^1_{all}-r^e_{all}$	and	n_f	-0.208	0.050

paper focuses only on the results of the ANOVA for VES, V2ES, and V2LS.

The results of the test of no correlation on the S_p based on the various quantities, show the correlations presented in the table. As a reference, for V1ES and V1LS, a correlation greater than $\rho = 0.5$ was observed.

IV. DISCUSSION:

Based on the $r^{1/2/e}_{all}$ and impression ratings obtained from the experimental results when presented stimuli with $t^i_p,\ t^i_{f,p}, T^i_{c,p},\ n^i_{p,f}$ and ANEW are given, we discuss the characteristics of the presented stimuli that are likely to be recalled.

A. Consideration of Observed Metrics

Table VI shows the medians of the presenting stimulus conditions in this paper, dT, the valence pair, and the presence or absence of recall and the type of S_p obtained from the experiment. The table does not include value pairs for which the results of the test of no correlation were not significant except for $(V_{--},\ V_{--}, dT = 7)$. First, focusing on $T^i_{c,p}$, the values are lower for $dT = 7,$ VES, $(V_{--},\ V_{++},\ dT = 7)$ and higher for V1ES, $(V_{++},\ V_{++},\ dT = 1)$, $(V_{--},\ V_{++},\ dT = 1)$. Next, t^i_p value is higher for V1ES, V1LS, and V2ES. The $t^i_{f,p}$ values are higher for V1ES, V1LS, and $(V_{++},\ V_{++},\ dT = 1)$, and lower for $(V_{--},\ V_{++},\ dT = 7)$. The total change in pupillary response is difficult to understand, but in general, the value is higher for V2LS and lower for the $r^{1/e}_{all}$ of $(V_{--},\ V_{++},\ dT = 7)$.

Based on the above consideration, we discuss the strong correlation with the total change in pupillary response for each value pair according to Table IV. The reason for this is that the degree of emotional induction to the presented stimulus is related to the degree of interference of multiple ANEWs, although emotional is induced by ANEWs, as assumed in Section II. When multiple ANEWs are presented to listeners during the interval of a dT, the emotions induced by the ANEWs may interfere with each other. If the dT is short enough to cause interference in emotion induction, then the emotions induced by the multiple ANEWs may interfere with each other, and the final emotion that remains in the listener's mind may induce more emotions than the original ANEWs induced, or it may repel the original ANEWs and not induce any emotions at all. The phenomenon that dT does not induce emotion is likely to occur. If the dT is long enough that it cannot cause interference in the emotion induction, then the emotion induced in the listener by multiple ANEWs will be the same as that induced by a single ANEW, or will induce only very weak interference.

In this study, the interference of induced emotion is considered to appear in the total change in pupillary response as shown in Figure 1. The quantities, $t^i_p,\ t^i_{f,p},\ n^i_{p,f}$, which occur in the participants' rating of emotion induction, rate the degree of emotion induced by the presented stimulus received by the listener as a valence, and the result is selected. Additionally, there may be some relationship between the selected S_p and $t^i_p,\ t^i_{f,p},\ n^i_{p,f}$. From this point of view, Tables V and IV show a clear correlation with the total change in pupillary response to the recalled presented stimulus with the valence pair of $(V_{--},\ V_{++},\ dT = 7)$, $(V_{++},\ V_{--},\ dT = 1)$. Moderate correlations were found for $(V_{--},\ V_{++},\ dT = 7)$ and $(V_{++},\ V_{--},\ dT = 1)$ to $r^1_{all} - r^e_{all}$ and r^2_{all}, respectively, with $t^i_{f,p}$ and $T^i_{c,p}$.

A comparison of this with the results from Table VI is as follows. $(V_{++},\ V_{--},\ dT = 1)$ does not show a trend of highly distorted $t^i_p,\ t^i_{f,p}$, and $n^i_{p,f}$ from the overall data. On the other hand, for $(V_{--},\ V_{++},\ dT = 7)$, $T^i_{c,p}$ is rather low,

TABLE VI. SUMMARY OF MEASUREMENT QUANTITIES.

		num. of data	num. of recall	ratio	$T^i_{c,p}$	t^i_p	$t^i_{f,p}$	r^1_{all}	r^2_{all}	r^e_{all}	$n^i_{p,f}$
all		544	161	0.30	0.7	3411.5	2220.33	1.92	2.02	1.53	7
dT	1	272	85	0.31	0.77	3416	2335	1.83	2	1.47	7
	7	272	76	0.28	*0.63*	3393	*1923.68*	2	2.12	1.57	6
recall	Yes	161		–	0.71	3436	2101.3	1.7	1.71	1.46	6
	No	380		–	0.69	3401	2236.29	1.99	2.2	1.64	7
type of evaluation	VES	149	52	0.35	*0.62*	3433	2232	1.85	2.02	1.5	6
	V1ES	10	1	0.10	**0.96**	**4880.5**	**3944.27**	*0.94*	1.94	*1.12*	9.5
	V1LS	19	3	0.16	0.72	**4240**	**3027.25**	1.72	3.27	1.7	7
	V2ES	117	35	0.30	0.79	3702	2645.18	1.92	1.79	1.41	7
	V2LS	90	25	0.28	0.68	3304.5	2004.99	**2.29**	2.03	**1.94**	6
valence pair	$(V_{++},\ V_{++},\ dT = 1)$	68	26	0.38	**0.95**	3646.5	**3123.6**	1.85	2.19	*1.39*	7
	$(V_{++},\ V_{++},\ dT = 7)$	68	14	0.21	0.73	3616.5	2072.66	2.09	2.31	1.71	6.5
	$(V_{++},\ V_{--},\ dT = 1)$	68	22	0.32	0.74	3149.5	2157.43	1.83	1.93	1.51	6.5
	$(V_{--},\ V_{++},\ dT = 1)$	68	19	0.28	**0.81**	3436	2335.55	1.89	1.55	1.43	6
	$(V_{--},\ V_{++},\ dT = 7)$	68	21	0.31	*0.64*	3617.5	*1890.56*	**2.26**	1.92	1.74	8
	$(V_{--},\ V_{--},\ dT = 7)$	68	24	0.35	*0.57*	3045.5	*1735.94*	1.96	1.98	1.47	*5*

$t^i_{f,p}$ is very low, and r^1_{all} is relatively higher. In addition, r^1_{all} shows a slightly larger change than the others. However, all of these valence pairs are comparable to each other in terms of the recall situation for the pair, as shown in Table III.

These results suggest that the pupil's response to a particular value pair and the response to recall may be somehow related to the pupil's response to a particular value pair.

B. Relation between Recall and Observed Metric at $dT = 1$

Table III shows the relationship between the number of recalls and the observed metrics. According to Table III, $(V_{++},\ V_{++})$ is the most frequently recalled, followed by $(V_{++},\ V_{--})$.

First, we focus on $(V_{++},\ V_{++})$. There is no correlation between the various quantities in Table IV. From Table VI, r^e_{all} shows lower values and $T^i_{c,p}$ shows higher values. The r^e_{all} indicates that the pupil diameter change was almost completely contained, suggesting that significant emotional induction remained. For $T^i_{c,p}$, the participants were gazing almost anywhere on the screen until S_p was determined. This suggests that $(V_{++},\ V_{++})$ rated S_p without significant emotional induction and their interference, but $T^i_{c,p}$ became longer as a result of hesitation over which to select when choosing S_p. This may be the reason why $T^i_{c,p}$ is longer than S_p. This may have led to cognitive load, and may have promoted recall due to the load.

Next, we focus on $(V_{++},\ V_{--})$. According to Table IV, the total change in pupillary response at the time of recall is correlated with r^2_{all} and $t^i_{f,p},\ n^i_{p,f}$. Considering the difference from the overall data, it shows a correlation between r^2_{all} and $t^i_{f,p}$, and checking the S_p values, 13 were classified as V2ES and five were set as V2LS, indicating that the answers were generally based on the second ANEW impression. According to Table VI, there are no areas that stand out compared to the overall trend. This suggests that the $(V_{++},\ V_{--})$ responses were based on the impressions received from the presented stimuli. In other words, the degree of cognitive load was low, suggesting that the impression was deep and thus facilitated recall.

C. Relation between recall and observed metric at $dT = 7$

We refer to the relationship between the number of recalls and the observed quantities shown in Table III. According to the table, $(V_{--},\ V_{--})$ is the most frequently recalled, followed by $(V_{--},\ V_{++})$.

First, we focus on $(V_{--},\ V_{--})$. Table IV shows no correlations or strong trends among the metrics. Table VI shows that $T^i_{c,p},\ t^i_{f,p},\ n^i_{p,f}$ both have low values. This indicates that the participants had almost no hesitation in determining S_p, and they checked the screen in a short amount of time. In other words, the degree of cognitive load is considered to be low, and may be due to the fact that the impression was deep, which facilitated recall.

Next, we focus on $(V_{--},\ V_{++})$. Table IV shows that the pupil diameter change at the time of recall is correlated with $r^1_{all} - r^e_{all}$ and $t^i_{f,p},\ T^i_{c,p}$. Since both have a common meaning, we can essentially assume that there is a correlation between $T^i_{f,p}$ and $r^1_{all} - r^e_{all}$. The S_p values for the recalled presented stimuli were eight for V2ES and seven for V2LS, indicating that the responses were generally based on the second ANEW impression. According to Table VI, $T^i_{c,p}, t^i_{f,p}$ are very low, while r^1_{all} is high. This suggests that a relatively high emotional induction occurred at the beginning of $(V_{--},\ V_{++})$, but the second ANEW impression was recalled independently of it, with that the difference from the emotional induction being high. This may be due to the cognitive load that facilitated the recall of the second ANEW impression.

D. Two Trends triggering recall

Based on the above, it can be concluded that there are several types of recall triggered by the presented stimulus with emotion induction.

- The degree of cognitive load due to emotion induction is low, but the stimulus is very impressive, which promotes recall.
- Promotion of recall due to a high degree of cognitive load induced by emotion induction

The former can be regarded as System 1, i.e., intuitive recall due to bias, and the latter as System 2, i.e., recall due to thought, as indicated by Kahneman [16].

In the case of a presented stimulus containing multiple ANEWs, it is necessary to consider emotion induction due to their interference with each other. It is not sufficient to prepare ANEW pairs at any time, nor is it sufficient to continuously prepare ANEW pairs of the same or different types. According to Table III, even for the same $(V_{++},\ V_{++})$ valence pair, $dT = 1$ and $dT = 7$ produce significantly different results for recall. Notably, the recall rate is almost halved for $dT = 7$ compared to $dT = 1$, suggesting a strong influence of interference generated during ANEW's emotion induction.

In this way, one possibility for a set of ANESs that promotes recall may be the valence pair and the placement of ANEWs within the range where they can interfere with each other. However, while many research institutes have made certain measurements of ANEW valence and can predict S_p evaluations based on these measurements, the interference between valences is not the same from person to person due to differences in individual responses to the valences and individual differences in the degree of interference between valences. The final response will differ from person to person. In the future, based on our findings, we believe that it will be possible to design content that is tailored to each individual after measuring the interference characteristics of the individual's emotion induction to the value pair using LGR-Map and other methods.

V. Conclusion and Future Works

In this paper, we analyzed the relationship between the valence of ANEWs in linguistic information and recall, using the total change in pupil diameter. We sought to understand the degree of emotional interference by the combination of valence pair and interval time between valence pairs for two ANEWs in short sentences, employing metrics such as the time required for impression evaluation, the total time spent during impression evaluation, the ratio of both, and the number of stopping points. Additionally, we also examined the effect of the combination of valence pair and interval time on the recall rate. The results showed that responses with a high recall rate were 1) The response with a high recall rate had a low cognitive load due to emotion induction, but was very impressive, which promoted recall. 2) The response with a high recall rate was associated with a high degree of cognitive load by emotion induction. These findings suggest the presence of two types of memories, namely System 1 and System 2, as described by Kahneman. Specifically, they correspond to bias-induced memory and thought-induced memory.

These characteristics may be intrinsic to each person, or they may be general properties that do not depend on the person. In the future, the development of a metric to discriminate between the two states of recall will be valuable for designing content that is easier to recall according to the individual.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 19K12246 / 19K12232 / 20H04290 / 22K12284 / 23K11334 , and National University Management Reform Promotion Project.

References

[1] B. Mahanama et al., "Eye movement and pupil measures: A review," Frontiers in Computer Science, vol. 3, 2022. [Online]. Available: https://par.nsf.gov/biblio/10313299

[2] A. Vilotijević and S. Mathôt, "Functional benefits of cognitively driven pupil-size changes," WIREs Cognitive Science, vol. 15, no. 3, 2024, p. e1672. [Online]. Available: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcs.1672

[3] V. Skaramagkas et al., "Review of eye tracking metrics involved in emotional and cognitive processes," IEEE Reviews in Biomedical Engineering, vol. PP, 03 2021, pp. 1–1.

[4] T. Partala and V. Surakka, "Pupil size variation as an indication of affective processing," International Journal of Human-Computer Studies, vol. 59, 07 2003, pp. 185–198.

[5] R. Henderson, M. Bradley, and P. Lang, "Emotional imagery and pupil diameter," Psychophysiology, vol. 55, 12 2017, p. e13050.

[6] M. Oliva and A. Anikin, "Pupil dilation reflects the time course of emotion recognition in human vocalizations," Scientific Reports, vol. 8, 03 2018.

[7] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. Rak, "Eye-tracking analysis for emotion recognition," Computational Intelligence and Neuroscience, vol. 2020, 09 2020, pp. 1–13.

[8] R. Hirabayashi, M. Shino, K. T. Nakahira, and M. Kitajima, "How auditory information presentation timings affect memory when watching omnidirectional movie with audio guide," in Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020), vol. 2, 2020, pp. 162–169.

[9] M. Murakami, M. Shino, K. T. Nakahira, and M. Kitajima, "Effects of emotion-induction words on memory of viewing visual stimuli with audio guide," in Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2021), vol. 2, 2021, pp. 89–100.

[10] K. T. Nakahira, M. Harada, and M. Kitajima, "Analysis of the relationship between subjective difficulty of a task and the efforts put into it using biometric information," in Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2022, Volume 2: HUCAPP, Online Streaming, February 6-8, 2022, A. Paljic, M. Ziat, and K. Bouatouch, Eds. SCITEPRESS, 2022, pp. 241–248. [Online]. Available: https://doi.org/10.5220/0010906800003124

[11] Y. Kurihara, M. Shino, K. Nakahira, and M. Kitajima, "Visual behavior based on information foraging theory toward designing of auditory information," in Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - HUCAPP, INSTICC. SciTePress, 2024, pp. 530–537.

[12] K. T. Nakahira, M. Harada, S. Moriya, and M. Kitajima, "Local-global reaction map for classifying listeners by pupil response to sentences with emotion induction words and its application to auditory information design," International Journal on Advances in Intelligent Systems, no. 1&2, 2024, pp. 38 – 48.

[13] W. Kintsch, "The use of knowledge in discourse processing: A construction-integration model," Psychological Review, vol. 95, 1988, pp. 163–182.

[14] W. Kintsch, Comprehension: A paradigm for cognition. Cambridge, UK: Cambridge University Press, 1998.

[15] C. Wharton and W. Kintsch, "An overview of construction-integration model: A theory of comprehension as a foundation for a new cognitive architecture," SIGART Bull., vol. 2, no. 4, jul 1991, p. 169–173. [Online]. Available: https://doi.org/10.1145/122344.122379

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