Language and Image in Behavioral Ecology

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, CNRS, INRIA, Loria
Nancy, France
Email: jerome.dinet@univ-lorraine.fr

Clelie Amiot
Univ. de Lorraine, CNRS, INRIA, Loria
Nancy, France
Email: clelie.amiot@univ-lorraine.fr

Capucine Bauchet
Univ. de Lorraine, DANE Nancy-Metz
Nancy, France
Email: capucine.bauchet@univ-lorraine.fr

Hanna Verdel
Univ. de Lorraine, DANE Nancy-Metz
Nancy, France
Email: hanna.verdel@univ-lorraine.fr

***Abstract*—Ideas are created in one's mind through cognitive processes after obtaining perceptual stimuli either by hearing or reading words or by seeing images. They should have different representations depending on their origin of information, i.e., words or images, and the cognitive processes for dealing with them. The comparison between these processes is often labeled by the terms, "word and wordless thought" and there is a strong argument that favors wordless thought. The purpose of this paper is to compare the two cognitive processes for words and images by applying the state of the art cognitive architecture, the Model Human Processor with Realtime Constraints (MHP/RT) proposed by Kitajima and Toyota, developed for understanding behavioral ecology of human beings. This study shows that the perceived dimensionality of images is larger than that of words, which leads to the conclusion that the number of discriminable states for images is an order of magnitude larger than that of words, and due to this, image-based processing can store information about absolute times in memory but word-based processing cannot. This should lend significantly larger expressive power to image-based processing. It is argued that the loss of reality in word-based processing results in significant implications for the development of globalization and the illusion of mutual understanding in word-level communications.**

Keywords–Word and wordless thought; Cognitive architecture; MHP/RT; Loss of reality.

I. Introduction

There is a teaching that says, "No matter how many times you *listen* to it, you can't actually *see* it even once. You should see everything with your own eyes." This is what Zhao Chongguo, a general of the Former Han, said as he introduced the following passage from a volume entitled "A History of Zhao Chongguo" in the Book of Han or History of Former Han:

> The Han Emperor asked Zhao Chongguo about the strategies and forces needed to quell the rebellious Tibetan nomads. Zhao Chongguo asked for forgiveness, saying, "Since it is difficult to formulate a strategy in a distant place, I would like to go to the site and draw a map of what I actually saw and tell a trick."

This somewhat abstract teaching is also demonstrated in the well-known sentence, "it is better to see it with your own eyes than to hear it a hundred times," which is an expression that can be translated into actions in a more understandable way. In this expression, hearing something a hundred times is equated with seeing it once. The following two translations are also derived from this: "a picture is worth ten thousand words" and "seeing is believing."

Regarding the first translation, Larkin and Simon [2] posed the following research question from the cognitive scientific viewpoint: "Why a diagram is (sometimes) worth ten thousand words." In order to find the answer, they assumed a situation where the same amount of information was expressed by diagrams and sentential paper-and-pencil representations and examined their characteristics from the viewpoint of human information processing to see if the amount of knowledge that one can acquire by *seeing a diagram once* is equivalent to the amount of knowledge that one can acquire by *hearing ten thousand words* that explain that diagram. Words and wordless thought are controversial issues in philosophy. One position identifies words or language with cognition, stating that an idea cannot be conceived other than through the word and only exists by the word, while wordless thought corresponds to the cognitive processes that are invoked when seeing a diagram. Jacques Salomon Hadamard, a distinguished French mathematician, described his own mathematical thinking as largely *wordless*, often accompanied by mental images that represent the entire solution to a problem. In his book entitled, "Psychology of Invention in the Mathematical Field [3]," he tried to report and interpret observations, either personal or gathered from other scholars engaged in the work of invention.

Regarding the second translation, "seeing is believing" implies that the state of believing something, which can be rephrased as the behavior of accepting the truth, reality, or validity of something (e.g., a phenomenon or a person's veracity), could be reached by seeing it. The perceptual and cognitive processes that occur between seeing the site of the rebellion and accepting the reality of the rebellion in the case of Zhao Chongguo, can be mapped on Two Minds [1][4], which suggests that human behavior emerges as the result of competition between the dual processes of System 2, which is a slow conscious process for deliberate reasoning with feedback control, and System 1, which is a fast unconscious process for intuitive reaction with feedforward control for connecting perception and motor. The section surrounded by the dotted line rectangle in Figure 1 is a modified version

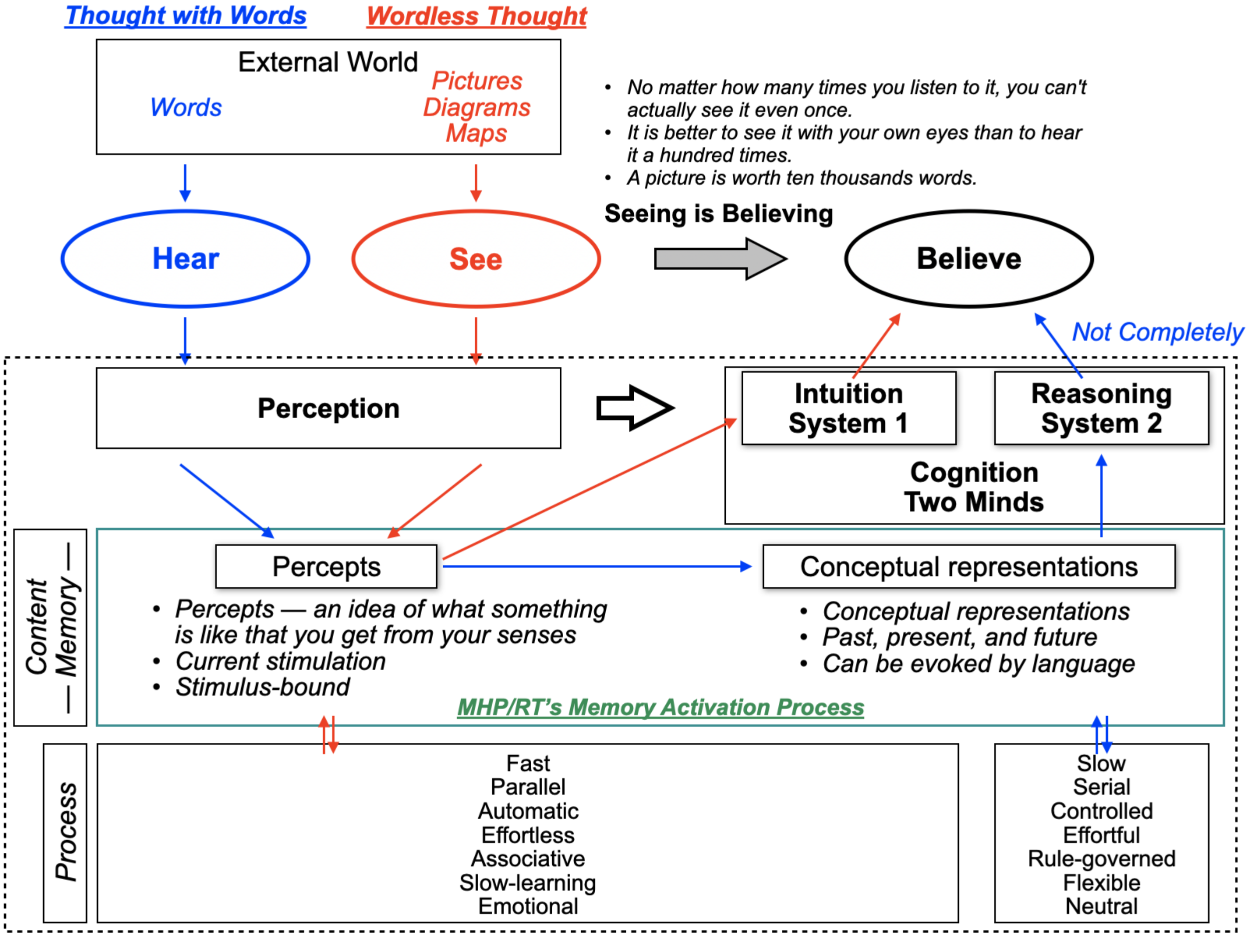

Thought with Words
Wordless Thought
External World
Words
Pictures
Diagrams
Maps
• No matter how many times you listen to it, you can't actually see it even once.
• It is better to see it with your own eyes than to hear it a hundred times.
• A picture is worth ten thousands words.
Seeing is Believing
Hear
See
Believe
Not Completely
Perception
Intuition System 1
Reasoning System 2
Cognition Two Minds
Content — Memory —
Percepts
• Percepts — an idea of what something is like that you get from your senses
• Current stimulation
• Stimulus-bound
Conceptual representations
• Conceptual representations
• Past, present, and future
• Can be evoked by language
MHP/RT's Memory Activation Process
Process
Fast
Parallel
Automatic
Effortless
Associative
Slow-learning
Emotional
Slow
Serial
Controlled
Effortful
Rule-governed
Flexible
Neutral

Figure 1. Word and wordless thought processes represented on the modified version of the figure used by Kahneman to explain Two Minds [1, Figure 2].

of the Figure used by Kahneman to explain Two Minds [1, Figure 2]. In Figure 1, the process that should correspond to “Seeing is Believing” is characterized as “Wordless Thought.” It starts by seeing pictures, diagrams, or maps provided in the external world, followed by forming a percept which is an idea of what something is like that you get from your senses and from intuitively reaching the state of believing through System 1. The behaviors expressed in the teachings referred to above, i.e., “hearing ten thousand words,” “to hear something a hundred times,” or “to listen to something many times,” are shown as “Thought with Words” in Figure 1. This process starts by hearing words, followed by forming a percept and reaching a state of not-completely-believing after extensive and deliberate reasoning processes through System 2.

In this paper, we aimed to dig deeper into the teaching in the Book of Han or History of Former Han and its variants by applying state-of-the-art cognitive architecture. In particular, the difference between Wordless Thought and Thought with Words will be clarified from the viewpoint of the difference in how layered structured memory is activated. The particular cognitive architecture we used for the analysis was the Model Human Processor with Realtime Constraints (MHP/RT) [5][6], which was applied to a variety of phenomena related to action selection and memory processes [7]–[15].

The remainder of this paper is organized as follows: Section II reviews MHP/RT focusing on the use of memories in the processing of System 1 and System 2, which is critical to understand the differences between image-based and word-based processing. Section III demonstrates that the distinction between word and wordless thought is one of the most ancient debates in psychology and has its roots in philosophical and educational considerations, and that several modern theories are based on this distinction. Section IV describes in detail how language and image are processed by MHP/RT focusing on how reality is guaranteed in these processes. Section V concludes the paper by summarizing the contents and pointing out the implications of the loss of reality in word-based communication in the development of globalization.

II. Model Human Processor with Realtime Constraints and Multi-dimensional Memory Frames

Kitajima and Toyota [6][16] constructed a comprehensive theory of action selection and memory, known as the Model Human Processor with Realtime Constraints (MHP/RT), that provides a basis for constructing any model for understanding human behavior (see Figure 2).

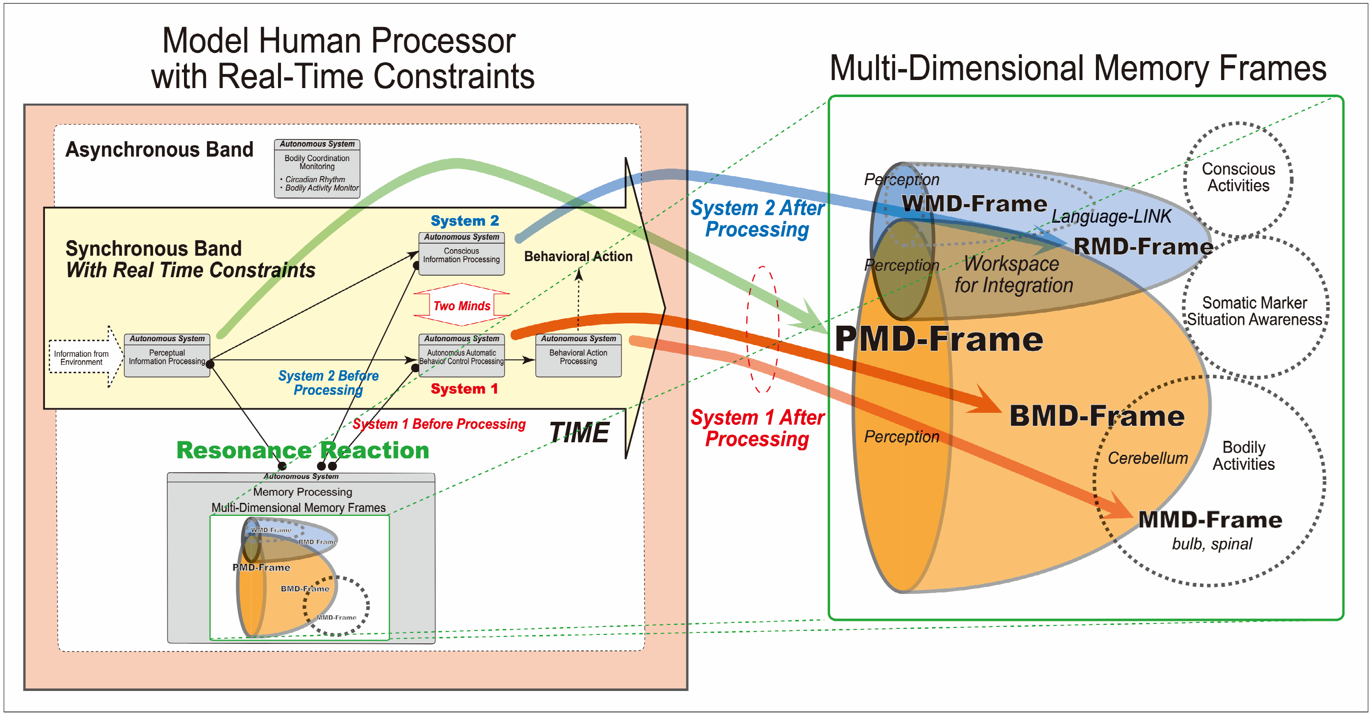

Model Human Processor
with Real-Time Constraints
Asynchronous Band
Autonomous System
Bodily Coordination Monitoring
• Circadian Rhythm
• Bodily Activity Monitor
Synchronous Band
With Real Time Constraints
System 2
Autonomous System
Conscious Information Processing
Behavioral Action
Two Minds
Information from Environment
Autonomous System
Perceptual Information Processing
Autonomous System
Autonomous Automatic Behavior Control Processing
Autonomous System
Behavioral Action Processing
System 2 Before Processing
System 1
System 1 Before Processing
TIME
Resonance Reaction
Autonomous System
Memory Processing
Multi-Dimensional Memory Frames
WMD-Frame
RMD Frame
PMD-Frame
BMD-Frame
MMD-Frame
Multi-Dimensional Memory Frames
System 2 After Processing
System 1 After Processing
Perception
WMD-Frame
Language-LINK
RMD-Frame
Perception
Workspace for Integration
Conscious Activities
Somatic Marker Situation Awareness
PMD-Frame
BMD-Frame
Perception
Cerebellum
Bodily Activities
MMD-Frame
bulb, spinal

Figure 2. MHP/RT and the distributed memory system implemented as multi-dimensional memory frames (modified from [6, Figure 3]).

A. Outline of MHP/RT

The MHP/RT is an extension of the Model Human Processor proposed by Card, Moran, and Newell [17] that can simulate routine goal-directed behaviors. The process involved in action selection is a dynamic interaction that evolves in the irreversible time dimension. The purpose of MHP/RT is to explain the following three facts that underpin an understanding of the behavioral ecology of human beings:

1) The fundamental processing mechanism of the brain is Parallel Distributed Processing (PDP) [18], which is referred to as the Organic PDP (O-PDP) system in the development of MHP/RT.
2) Human behavior emerges as a result of competition between the dual processes of System 1, fast unconscious processes for intuitive reaction with feed-forward control that connect perception with motor movements, and System 2, slow conscious processes for deliberate reasoning with feedback control. This is called Two Minds [1].
3) Human behavior is organized into 17 happiness goals [19].

B. Part 1: PCM Processes

MHP/RT consists of two parts. The first comprises cyclic PCM processes (Figure 2, left), in which PDP for these processes is implemented in hierarchically organized bands with characteristic operation times by associating relative times (not absolute) with the PCM processes that carry out a series of events that are synchronous with changes in the external environment. There is a gap between two adjacent bands; these two bands are non-linearly connected and therefore it is inappropriate to understand the phenomena that occur across these bands by constructing a linear model. The phenomena occur by connecting what happens in a band to what happens in its adjacent band non-linearly. A mechanism is required to connect the phenomena; MHP/RT suggests that this connection is provided by the resonance mechanism via the multi-dimensional memory frames.

C. Part 2: Multi-Dimensional Memory Frames

The bottom-left and right sections of Figure 2 show the autonomous memory system consisting of multi-dimensional memory frames of perception, motion, behavior, relation, and word. These memory frames store information associated with the corresponding autonomous processes defined in the PCM processes. The memory frames are subservient to the PCM processes because they do not exist unless the PCM processes exist.

The right section of Figure 2 shows the five memory frames and their relationship with the PCM processes. The following provides brief explanations of the respective memory frames.

- **WMD (Word MD)-frame** is the memory structure for language. It is constructed on a very simple one-dimensional array.
- **RMD (Relation MD)-frame** is the memory structure associated with the conscious information processing. It combines a set of BMD-frames into a manipulable unit.
- **BMD (Behavior MD)-frame** is the memory structure associated with the autonomous automatic behavior

control processing. It combines a set of MMD-frames into a manipulable unit.

- **PMD (Perceptual MD)-frame** constitutes perceptual memory as a relational matrix structure. It incrementally grows as it creates memory from the input information and matches it against the past memory in parallel.
- **MMD (Motion MD)-frame** constitutes behavioral memory as a matrix structure. It gathers a variety of perceptual information as well to connect muscles with nerves using spinals as a reflection point. In accordance with one's physical growth, it widens the range of activities the behavioral action processing can cover autonomously.

The memory frames have overlapping regions as follows: the PMD-frame overlaps with the WMD-, the RMD-, and the BMD-frames; the WMD-frame overlaps with the RMD-frame; and the BMD-frame overlaps with the MMD-frame. The PCM processes work for carrying out appropriate actions in response to the input stimuli. These actions are carried out when the corresponding portions of the MMD-frames are activated. The MMD-frames only overlap with the BMD-frame, and not with the RMD-frame or the WMD-frame, which System 2 operates with. This means that there is no direct path for System 2 to the MMD-frame to initiate any actions. System 2 can only indirectly contribute to the real actions via the BMD-frame which is connected to the MMD-frame.

D. Resonance as a Mechanism for Interaction Between PCM Processes and Memories

An important feature of the memory system is that it works *asynchronously* with the external environment. MHP/RT assumes that the *synchronous* PCM processes, including the perceptual system, System 1, System 2, the motor system, and the asynchronous memory system communicate with each other through a resonance mechanism. The concept of resonance has been borrowed from physics to describe the link between the asynchronous memory system and synchronous PCM processes. As Dinet et al. [12] suggested, apprehension of psychological phenomena using concepts borrowed from physics is useful because the majority of the interactions, including psychological interactions, between humans and the environment (social or physical environment) can be derived from physical processes.

Through the resonance process, the memory-frames work for the PCM processes to map perceptual information represented in the dimensionality of M to a motion represented in the dimensionality of N. In other words, the memory frames implement the $M \otimes N$ mapping from perception to motion via the resonance mechanism that connects the memories with the PCM processes. Figure 3 presents the relationships between the PCM processes shown at the bottom and the multi-dimensional memory frames shown at the top of the figure. It is important to note that System 1 has direct and parallel paths from perception to motion via the PMD-, the BMD-, and the MMD-frames, whereas System 2 does not in the $M \otimes N$ mapping. System 2 carries out serial processing along the paths from the PMD-frame to the WMD- and the RMD-frames, which are the memory for System 2. The results of System 2's processing could be transferred to the MMD-frame that enables actual actions through the overlaps between the RMD-frame and the BMD-frame. This can be described as follows: there are *direct* mappings of $M \otimes N$ for System 1 and there are *indirect* contributions of System 2's workings in these mappings, which is informally denoted as $M \otimes$(WORDS)$\otimes N$.

III. Word and Wordless Thought

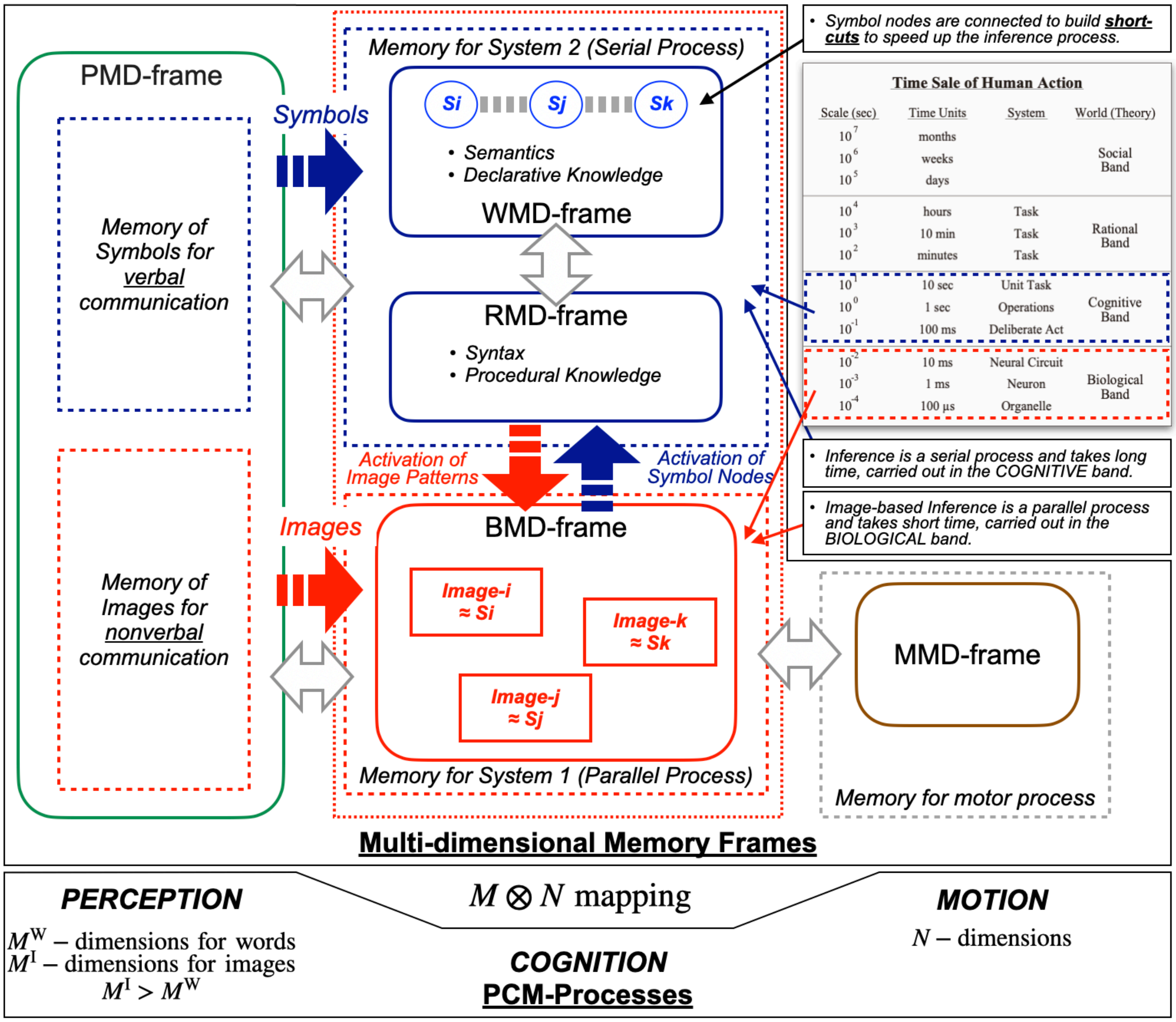
This section aims to demonstrate that (i) the distinction between word and wordless thought is one of the most ancient debates in psychology and has its roots in philosophical and educational considerations, and (ii) several modern theories (e.g., Dual Coding Theory and Cognitive Theory of Multimedia Learning) are based on this distinction.

A. The Philosophical Roots of the Debate

The distinction between word and wordless thought is believed to have inspired all cognitive and behavioral sciences that are interested in all forms of psychological explanations for the behavior of non-linguistic and linguistic creatures [20].

This distinction between word and wordless thought has been one of the bases of all educational manuals for several centuries. For instance, the educational pioneer, Jan Amos Komensky (alias Comenius), was the first author who, in the 17th century, proposed manuals that combined texts and pictures. His book entitled "Orbis Sensualium Pictus" (translated as "The world explained in pictures"), published in 1658 in Nuremberg, was the first widely used children's textbook with pictures. It was first published in Latin and German and later republished in many European languages, quickly spreading around Europe and becoming the definitive children's textbook for three centuries, with more than 200 editions published in twenty-six languages. This manual contains an extended summary of the world in 150 pictures with titles (Figure 4, left). All of the objects in the pictures are numbered and accompanied by parallel columns of labels and short sentences describing the numbered objects, categorized in different domains (zoology, religion, botany, etc.). More than 350 years later, as the right-hand section of Figure 4 shows, education manuals always have a similar appearance, where text and images are combined, because it is assumed that this combination has a positive impact on understanding and memory of information [21][22]. The main difference between manuals printed in the 17th century and modern manuals is that colors and typographical cues have been added due to progress in modern printing.

While some authors have been proposing materials that combine word and wordless thought for several centuries, psychological theories that explain the impact of this combination are more recent. For instance, Dual Coding Theory (DCT) has its roots in the practical use of imagery as a memory aid dating back 2,500 years [23][24]. Cognition, according to DCT, involves the activity of two distinct subsystems: (i) a verbal system specialized for dealing directly with language; and (ii) a non-verbal (imagery) system specialized for dealing with non-linguistic objects and events. The systems are assumed to be composed of internal representational units, called logogens and imagens, that are activated when one recognizes,


PMD-frame
Memory of Symbols for verbal communication
Memory of Images for nonverbal communication
Symbols
Images
Memory for System 2 (Serial Process)
Si
Sj
Sk
• Semantics
• Declarative Knowledge
WMD-frame
RMD-frame
• Syntax
• Procedural Knowledge
Activation of Image Patterns
Activation of Symbol Nodes
BMD-frame
Image-i ≈ Si
Image-k ≈ Sk
Image-j ≈ Sj
Memory for System 1 (Parallel Process)
MMD-frame
Memory for motor process
Multi-dimensional Memory Frames
PERCEPTION
M^W – dimensions for words
M^I – dimensions for images
$M^I > M^W$
$M \otimes N$ mapping
COGNITION
PCM-Processes
MOTION
N – dimensions

- Symbol nodes are connected to build ***short-cuts*** to speed up the inference process.

Time Sale of Human Action

Scale (sec)	Time Units	System	World (Theory)
10^7	months		Social Band
10^6	weeks		
10^5	days		
10^4	hours	Task	Rational Band
10^3	10 min	Task	
10^2	minutes	Task	
10^1	10 sec	Unit Task	Cognitive Band
10^0	1 sec	Operations	
10^{-1}	100 ms	Deliberate Act	
10^{-2}	10 ms	Neural Circuit	Biological Band
10^{-3}	1 ms	Neuron	
10^{-4}	100 µs	Organelle	

- *Inference is a serial process and takes long time, carried out in the COGNITIVE band.*
- *Image-based Inference is a parallel process and takes short time, carried out in the BIOLOGICAL band.*

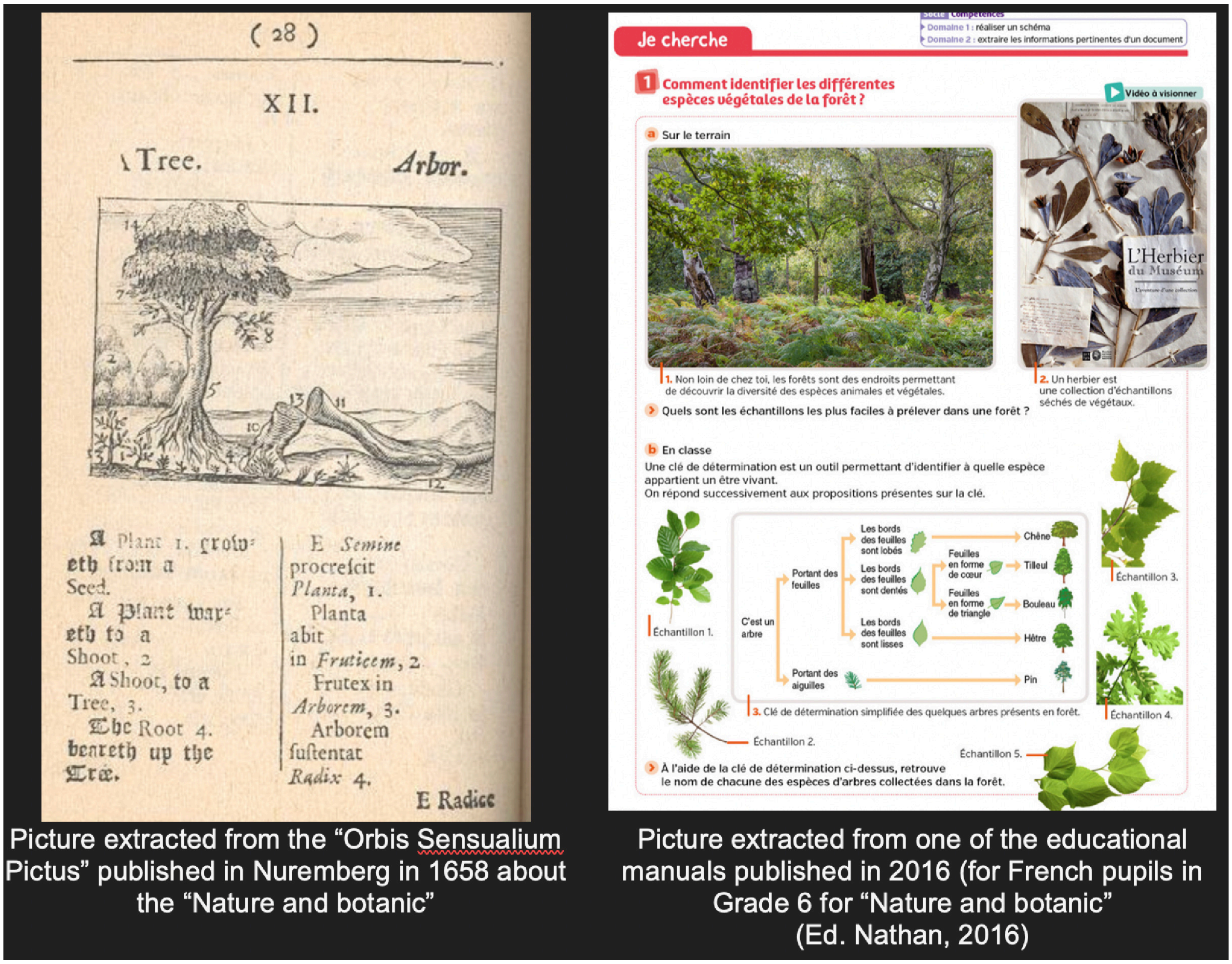
Figure 3. The relationships between the PCM processes shown in the bottom of Figure 3 and the multi-dimensional memory frames shown in the top of Figure 3.

manipulates, or just thinks about words or things. From a developmental point of view, dual coding development begins with the formation of a substrate of non-verbal representations and imagery derived from a child's observations and behaviors related to concrete objects and events, and relations among them. Language builds upon this foundation and remains functionally connected to it as referential connections are being formed, so that the child responds to object names in the presence or absence of the objects and begins to name and describe them (even in their absence). The events, relations, and behaviors are dynamically organized (repeated with variations) and thereby display natural syntax that is incorporated into the imagery as well. The natural syntax is enriched by motor components derived from the child's actions, which have their own patterns.

B. Modern Psychological Theories

A series of behavioral and neuropsychological studies provide further relevant support for this dual-channel approach (word versus wordless thought). For example, Thompson and Paivio [25] showed that object pictures and sounds had additive effects on memory, thereby supporting the DCT assumption that sensory components of multimodal objects are functionally independent. Similar effects have been demonstrated for combinations of other modalities. Brain scan studies have shown that different brain areas are activated by concrete and abstract words, as well as by pictures, as compared to words in comprehension and memory tasks (summarized in [23]). Other meta-analyses examined the most common loci of activation in fMRI and PET studies comparing abstract and concrete conceptual representations and support the dual-channel approach [26][27].

More recently, other theories related to psychology and education sciences have also been based on the distinction between word and wordless thought. For instance, the Cognitive Theory of Multimedia Learning, which is based on the Cognitive Load Theory and DCT, was developed after considering the previous theories, and is defined as the learning that



(28)

XII.

A Tree. Arbor.

A Plant 1. groweth from a Seed.
A Plant waxeth to a Shoot, 2
A Shoot, to a Tree, 3.
The Root 4. beareth up the Tree.

E Semine procreſcit Planta, 1.
Planta abit in Fruticem, 2
Frutex in Arborem, 3.
Arborem ſuſtentat Radix 4.

E Radice

Picture extracted from the "Orbis Sensualium Pictus" published in Nuremberg in 1658 about the "Nature and botanic"

Je cherche

Socle Compétences
- Domaine 1 : réaliser un schéma
- Domaine 2 : extraire les informations pertinentes d'un document

1 Comment identifier les différentes espèces végétales de la forêt ?

Vidéo à visionner

a Sur le terrain

1. Non loin de chez toi, les forêts sont des endroits permettant de découvrir la diversité des espèces animales et végétales.

2. Un herbier est une collection d'échantillons séchés de végétaux.

› Quels sont les échantillons les plus faciles à prélever dans une forêt ?

b En classe

Une clé de détermination est un outil permettant d'identifier à quelle espèce appartient un être vivant.
On répond successivement aux propositions présentes sur la clé.

C'est un arbre
Portant des feuilles
Les bords des feuilles sont lobés
Chêne
Les bords des feuilles sont dentés
Feuilles en forme de cœur
Tilleul
Feuilles en forme de triangle
Bouleau
Les bords des feuilles sont lisses
Hêtre
Portant des aiguilles
Pin
Échantillon 1.
Échantillon 2.
Échantillon 3.
Échantillon 4.
Échantillon 5.

3. Clé de détermination simplifiée des quelques arbres présents en forêt.

› À l'aide de la clé de détermination ci-dessus, retrouve le nom de chacune des espèces d'arbres collectées dans la forêt.

Picture extracted from one of the educational manuals published in 2016 (for French pupils in Grade 6 for "Nature and botanic" (Ed. Nathan, 2016)

Figure 4. Educational manuals through the time, from the 17th Century to 21st Century.

is realized when constructing mental representations through pictures and words [21][28]–[30]. The Cognitive Theory of Multimedia Learning addresses how individuals process information and how they learn through multimedia approaches. It encompasses three fundamental assumptions:

1) People have two separate channels for processing visual and audio information.
2) Each channel has a limited amount of information per unit of time.
3) People experience active learning by accessing related information, organizing the selected information through mental structures, and integrating them with previous mental structures.

In other words, the dual-channel assumption is incorporated into the Cognitive Theory of Multimedia Learning by proposing that the human information-processing system contains an auditory/verbal channel and a visual/pictorial channel. According to Mayer [21][28]–[30], the relationship between the two channels is under the control of the user in the sense that users may be able to convert the representation for processing in the other channel if their cognitive capacity is sufficient. The limited capacity assumption (i.e., humans are limited as to the amount of information that can be processed in each channel at one time) is also very important for the Cognitive Theory of Multimedia Learning. Metacognitive strategies are techniques for allocating and adjusting the limited cognitive resources of the center executive (i.e., the system that controls the allocation of cognitive resources) and play a central role in all modern theories of intelligence [31].

C. Debate is Always Actual

Finally, we note that the debates between word and wordless language are at the core of recent studies investigating production and comprehension of a new generation of emoticons that have continued to grow in popularity and usage in both mobile communications and social media [32]. More precisely, because emojis have a distinctively social nature and are arguably extra-linguistic in origin, in that they are not part of the standard lexicon of any natural language, they have a special place in digital communication [33]. Today, emojis are pictographs that have become common units of expression. They are non-verbal cues for emotion (anger, joy, celebration) and similarly evocative expressions in digital communication, including text messages sent through smartphones and on social media platforms such as Twitter [32].

Copyright (c) IARIA, 2023. ISBN: 978-1-61208-950-8

IV. Language and Image Processing in Behavioral Ecology

As an example of a situation where the teaching "seeing is believing" can be applied, we can use the sight of cherry blossoms falling. When a person actually sees this scene, s/he perceives it as a visual image. When s/he is provided with the description of the scene, e.g., "the beauty of cherry blossoms fascinates people, not only because of the blossoms themselves, but also because they are short lived" via printed or handwritten text on paper, or as an audible voice from a speaker or a person, s/he perceives it verbally as visual or auditory language. The perceived information is stored in working memory as cognitive frames for further processing [5, Figure 5]. Three things could follow, as indicated by the three lines leaving the box labeled "Perceptual Information Processing" in Figure 2:

1) The perceived information resonates with the multi-dimensional memory frames.
2) It is transferred to System 2 to initiate deliberate and conscious processing, e.g., inferences.
3) It is transferred to System 1 to initiate automatic and unconscious processing such as immediate emotional reactions for the voices.

While System 2 and System 1 are processing information, they can incorporate the activated portion of the memories through resonance. These processes are shown graphically by ●—● in Figure 2 by connecting System 2 and System 1 with multi-dimensional memory frames. This section focuses on how MHP/RT uses memory to process languages and images in order to clarify the differences between these processes.

A. Processing Languages

This section focuses on how verbal inputs are processed by the PCM processes and the multi-dimensional memory frames.

1) Transferring Perceived Words to the WMD-frame: The perceived information presented verbally could resonate with the contents in the PMD-frame, i.e., the memory of symbols for verbal communication as shown at the top of the PMD-frame in Figure 3. The resonance would spread within the PMD-frame and transfer to the WMD-frame, where semantics of concepts is stored as declarative knowledge. The WMD-frame overlaps with the RMD-frame. These memory frames jointly advance language processing.

The following section examines in detail what would happen in the WMD-frame after the perceived information has arrived. Firstly, the structure of the WMD-frame, which is *functionally* constructed on a very simple one-dimensional array [34], should impose a strong restriction on the results of the resonance between the perceived information and the PMD-frame. Let L_i $(i = 1, \cdots)$ be a node in the WMD-frame and L_α be the node that receives the information from the PMD-frame. Since the memory related with L_α is constructed as a one-dimensional array, the activations originated from L_α could spread to $\{(L_{\alpha-1},\ L_{\alpha+1}),\ (L_{\alpha-2},\ L_{\alpha+2}),\ \cdots,\ (L_{\alpha-n},\ L_{\alpha+n})\}$ as time goes by along the connected nodes. After the PMD-frame establishes resonance with the perceived information, activation could spread in the WMD-frame with the overlapping node, L_α, as its origin.

2) Structure of the WMD-frame: The WMD-frame stores two types of language. The first type is "spoken language," i.e., *parole*, which defines spontaneously generated repetitive usage patterns of *phonetic symbols*. It is used as a means for exchanging information in the collective ecologies of humans. It should mirror human–human bodily interactions that are carried out under the structure of bodily functions [35]. The more frequently a sequence of phonetic symbols is used, the tighter it becomes to form a firm one-dimensional array of phonetic symbols. Possible interactions are restricted by the context in which they occur, which then restricts the range of spoken language that could emerge. This leads to construction of a thesaurus, which is a form of controlled vocabulary that seeks to dictate semantic manifestations of spoken language, and to make use of them while interacting with others. Semantic information could be hierarchically organized according to the levels of abstraction as information accumulates.

The second type is "written language," i.e., *langue*, which is a notational system with *logical symbols*, created by thoughtfully and evolutionarily developing spoken language for the purpose of efficient communication. The strong generality found in written languages leads to the establishment of grammar, i.e., *syntax*. The syntactic rules are stored in the RMD-frame, which could be activated by words stored in the WMD-frame using the overlap between them. This defines how System 2 carries out inference by utilizing the WMD-frame for semantics and the RMD-frame for syntax, which is a serial process and takes a long time, carried out in the cognitive band of Newell's time scale of human action [36, Fig. 3-3]. The one-dimensionality of the WMD-frame is the direct reflection of the nature of System 2's processing, which is serial processing. This means that only one node could be focused while performing inference. Therefore, the trajectory of inference processes could be represented as a series of one-by-one focused-on nodes.

B. Language vs. Image in Expressing Reality

1) Language Loses the Information Concerning Absolute Times: As described in Section IV-A, language realizes efficient human–human communication by generalizing repetitive usage patterns of phonetic or logical symbols. The process of generalization is called abstraction from the viewpoint of the degree of concreteness, or simplification from the viewpoint of the level of detail in expressing the situation. A generalized pattern represents wide variations of its instantiations in the real world. It can be used in human–human communication to express specific instantiations by using the generalized pattern that implies them.

As such, language is not appropriate for expressing reality, which is a collection of concrete instantiations in the real world. This is because language solely stores relationships, $\texttt{Rel}$, between events $E(T_1)$ and $E'(T_2)$ that happen at times T_1 and T_2, respectively. Let us suppose that $E(T_1)$ is an instance of a set of events $\boldsymbol{E}$ and $E'(T_2)$ is an instance of a set of events $\boldsymbol{E'}$. Language expresses this by $\texttt{Rel}(E(\cdot), E'(\cdot))$ which can be read as follows: the event $E(\cdot)$ is related to the event $E'(\cdot)$ by the relation $\texttt{Rel}$. In this way, when storing the relationship between the events, language loses the information of the absolute times T_1 and T_2 and accomplishes abstraction of concrete events. This means that it is inherently impossible for language to restore the lost information about the absolute times. $\texttt{Rel}(E(\cdot), E'(\cdot))$ is stored in the RMD-frame and $E(\cdot)$ and $E'(\cdot)$ are stored in the WMD-frame.

2) Image Can Maintain Reality: The loss of the information concerning absolute times is clearly understood by using a class of words for expressing movement, e.g., *take*, *make*, *get*, and *do*, as an example. They do not hold precise information in the time dimension as they are stored in the WMD-frame.

What happens precisely in the real world when we say “take some flowers to the hospital” is, e.g., 1) buy flowers at 9:00 a.m. at a flower shop; 2) arrive at the hospital at 9:30 a.m. The contents in the WMD-frame relevant to this event are contrasted with those in the PMD-frame. Information stored in the PMD-frame maintains the information about the absolute times. Movement of an object in the real world is defined precisely as a trajectory represented by a time series of the locations of the object in the three-dimensional space. Perceptual information associated with the movement of the object is processed by respective sensory organs at their characteristic sensing rates, e.g., visual information is processed at the rate of 100 msec per characteristic event. The perceptual information may resonate with the contents in the PMD-frame which is an accumulation of the past perceptual experiences for confirming their existence in the PMD-frame.

3) Recovery of Lost Information: The lost information concerning absolute times of events could be recovered with the help of the information stored in the BMD-frame, where the compiled motion patterns for the specific perceptual images are stored without the loss of the information of absolute times, which are critical for producing motion sequences synchronous with the external environment. The BMD-frame is used in System 1 for carrying out image-based inference, which operates at the biological band of the time scale of human action [36, Fig. 3-3] at the time range of < 100 msec.

Imagine you heard the sentence: “My husband brought flowers to the hospital.” This verbal information would resonate with the PMD-frame and then activate the event nodes in the WMD-frame that represent relevant events such as “My husband bought flowers,” “My husband went to the hospital,” and so on, after a series of inferences by applying procedural knowledge stored in the RMD-frame. These nodes are represented as a connected pattern of symbols in the WMD-frame; they are represented by S_i and S_j in Figure 3. The overlap between the WMD- and RMD-frames, and the BMD-frame could activate the nodes in the BMD-frame, which are the compiled motion patterns for the specific perceptual images. S_i and S_j in the WMD-frame could be associated with *Image-i* and *Image-j*, respectively, which might or might not be close to the sight S_i represents. There is no guarantee it is the real image, but the absolute time is recovered with no guarantee of its reality. In this case, the image nodes might represent the motion patterns for the husband’s flower-taking and going-to-hospital events, that have been individually encoded at any time in the past.

The use of the information in the BMD-frame has another advantage for the inference processes carried out in System 2. Suppose that *Image-k* follows *Image-j* in the BMD-frame. This could activate the symbol node S_k in the WMD-frame, which results in the establishment of the pattern of connection, $S_i \Rightarrow S_j \Rightarrow S_k$ in the WMD-frame. When the results of inferences are guaranteed by the time-guaranteed BMD-frame, even if there is no guarantee of its reality, they could be used as a shortcut to speed up the inference process. Retrieving images in the BMD-frame through the resonance mechanism would require a lot of effort, which a person would want to avoid. Once these connections are established, they are used as if it is possible to recover the lost information without making an effort to confirm its reality.

4) Reality in Self Experience: Consider another situation where you told a friend, “I took the flowers to the hospital.” This describes your own experience and differs from the previous example, which describes the behavior of one’s husband. In this case, the utterances heard via your ears would resonate with your PMD-frame followed by spreading activation to the words stored in the WMD-frame, to the procedural rules stored in the RMD-frame, and finally to the encoded sequence of behaviors stored in the BMD- and the MMD-frames, where concrete time series of the behavior just mentioned would be reproduced. In other words, when a person tells his/her own experience to someone else, its reality would be assured in himself or herself; while a person who heard someone else’s experience would never reproduce its reality, i.e., recovery of the lost information could not be accomplished completely when language is used for human–human communication.

5) Supporting Reality: How would MHP/RT and the multi-dimensional memory frames function when the input stimuli are images or non-words? As shown in Section II, the multi-dimensional memory frames help the PCM processes map the M-dimensional sensory stimuli onto the N-dimensional motions. The following demonstrates the richness of non-verbal image-based communication compared with verbal word-based communication, which should support the reality.

Let the dimensionality of images be M^{I} and that of words be M^{W}. M^{I} is the number of dimensions for *non-verbal* communication. Let M_i^{I} be the i-th dimension of non-verbal communication and N_i^{I} be the number of discriminable states for M_i^{I}. The number of discriminable states in non-verbal communication, N^{I}, is $\prod_{i=1}^{M^{\mathrm{I}}} N_i^{\mathrm{I}}$. M^{W} is the number of dimensions used for *verbal* communication which corresponds to the number of categories used for representing symbols, such as alphabets, in verbal communication, N^{W}, is $\prod_{j=1}^{M^{\mathrm{W}}} N_j^{\mathrm{W}}$. It would be reasonable to assume that M^{W} is smaller than M^{I}, $M^{\mathrm{W}} < M^{\mathrm{I}}$, and the numbers of discriminable states for any categories of words, M_j^{W}, are smaller than those for any M_i^{I}’s. Therefore, the number of discriminable states differs by an order of magnitude difference in power between the non-verbal space composed of perception and the verbal space composed of the symbols $N^{\mathrm{I}} \ggg N^{\mathrm{W}}$.

6) Seeing is Believing: Finally, consider a situation where no language appears. For example, you feel ephemeral when you are observing the cherry blossoms falling in front of you. The visual event resonates with the memories stored in the PMD-frame, which would activate images in the BMD-frame overlapping with the memory for motor processes stored in the MMD-frame. In other words, the visual event that is occurring in front of you activates the memory traces stored in the BMD- and MMD-frames to help to imagine a variety of situations that could occur next in reality. Symbol nodes in the WMD-frame would be activated as shown in the upward arrow from the BMD-frame to the RMD- and WMD-frames in Figure 3 for verbally explaining the situation and for deriving the reasons for the event, which leads to an acceptance that the statement is true or that something exists, i.e., the state of believing.

V. Conclusion

This paper discussed the differences between the inference processes initiated by language and image. There are teachings that suggest that image-based inference should be superior to language-based inference, such as "seeing is believing" and "a picture is worth ten thousand words." The experience of seeing or hearing is perceptual. This is transferred to cognitive and motor processes to act appropriately in the external environment. The crucial difference between language and image is the size of the dimensionality when they are perceived. It was suggested that the fact that the dimensionality of image M^{I} is larger than that of language M^{I} should lead to huge differences in the number of discriminable states at the stage of perception. Accordingly, the language input is processed serially in System 2 and the image input is processed in parallel in System 1. The former is a slow, deliberate, and rational process and the latter is a fast, automatic, and instinctive process.

It was suggested that the low expressive power and the slow processing speed in language processing should trade the reality of the represented concept for the rich expression of the concept that includes the information of absolute times attached to the respective concepts. As the information of absolute times is necessary to reproduce the event precisely, representation of events in terms of language is inherently impossible for reproducing the event but it can do so approximately, at best. In the world of language that has lost time, even if reality can almost be restored, human beings tend to use their experience to secure reality and stop making efforts to restore it from the next event onwards.

It was shown that the inherent inability to recover reality is the primary reason for the suggested implication of superiority of images over language in comprehending the situations represented by them from the analysis obtained by applying state-of-the-art cognitive architecture. An image belongs to an individual, which enables him/her to carry out inferences guaranteed by reality. However, words are shared by the public as symbols to communicate one's thoughts. Words might be associated with the patterns of images stored in the BMD-frame by following the memories from the WMD-frame where the words are stored to the BMD-frame via the RMD-frame. In this way, the public symbols in the WMD-frame are indirectly and approximately associated with the individual image patterns in the BMD-frame, which might be individually different. Language-based inference is based on the accumulation of knowledge up to that point, i.e., a meme [15], so it is strongly bound by it and cannot deviate from its scope. Conversely, images form a perception away from language, which makes it easier to approach the truth.

The differences in the image patterns that individuals would associate with a single symbol might be small when they belong to a single culture. This is one of the conditions for words to work as memes [37] where symbols used in a culture are associated with shared image patterns that guarantee reality. In other words, when the same word is used in a communication between individuals from different cultures, there is there is no guarantee that the shared image will correspond with reality. Since a meme is a common interpretation that is valid within the group to which each individual belongs, it differs between groups. The person from one culture who uses the words to communicate his/her thoughts tends to believe that the images associated with the words should be communicated as effectively to the other person irrespective of the culture s/he is from. If the other person is from a different culture, there is little chance of coincidence in the associated images that these two people will infer from the words. The inherent discrepancies in the reality associated with commonly used words will become problematic for globalization.

Acknowledgment

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

References

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

[2] J. H. Larkin and H. A. Simon, "Why a diagram is (sometimes) worth ten thousand words," Cognitive Science, vol. 11, no. 1, 1987, pp. 65–100.

[3] J. Hadamard, An Essay on the Psychology of Invention in the Mathematical Field. New York, NY, USA: Dover Publications, 1954.

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

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

[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 and M. Toyota, "Two Minds and Emotion," in COGNITIVE 2015 : The Seventh International Conference on Advanced Cognitive Technologies and Applications, 2015, pp. 8–16.

[8] M. Kitajima, S. Shimizu, and K. T. Nakahira, "Creating memorable experiences in virtual reality: Theory of its processes and preliminary eye-tracking study using omnidirectional movies with audio-guide," in 2017 3rd IEEE International Conference on Cybernetics (CYBCONF), June 2017, pp. 1–8.

[9] M. Kitajima, "Nourishing problem solving skills by performing hci tasks – relationships between the methods of problem solving (retrieval, discovery, or search) and the kinds of acquired problem solving skills," in Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018), vol. 2: HUCAPP. Setúbal, Portugal: SCITEPRESS – Science and Technology Publications, 2018, pp. 132–139.

[10] M. Kitajima, "Cognitive Chrono-Ethnography (CCE): A Behavioral Study Methodology Underpinned by the Cognitive Architecture, MHP/RT," in Proceedings of the 41st Annual Conference of the Cognitive Science Society. Cognitive Science Society, 2019, pp. 55–56.

[11] 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.

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

[13] J. Dinet and M. Kitajima, “Immersive interfaces for engagement and learning: Cognitive implications,” in Proceedings of the 2015 Virtual Reality International Conference, ser. VRIC ’18. New York, NY, USA: ACM, 2018, pp. 18/04:1–18/04:8. [Online]. Available: https://doi.org/10.1145/3234253.3234301

[14] M. Kitajima, “Cognitive Science Approach to Achieve SDGs,” in COGNITIVE 2020 : The Twelfth International Conference on Advanced Cognitive Technologies and Applications, 2020, pp. 55–61.

[15] M. Kitajima, M. Toyota, and J. Dinet, “The Role of Resonance in the Development and Propagation of Memes,” in COGNITIVE 2021 : The Thirteenth International Conference on Advanced Cognitive Technologies and Applications, 2021, pp. 28–36.

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

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

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

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

[20] J. L. Bermúdez, Thinking without words. Oxford University Press, 2007.

[21] R. E. Mayer, “Research-based principles for the design of instructional messages: The case of multimedia explanations,” Document Design, vol. 1, no. 1, 1999, pp. 7–19. [Online]. Available: https://www.jbe-platform.com/content/journals/10.1075/dd.1.1.02may

[22] R. E. Mayer, “Multimedia learning,” in Psychology of learning and motivation. Elsevier, 2002, vol. 41, pp. 85–139.

[23] A. Paivio, “Dual coding theory and education,” in Draft chapter presented at the conference on Pathways to Literacy Achievement for High Poverty Children at The University of Michigan School of Education, 2006.

[24] M. Sadoski and A. Paivio, Imagery and text: A dual coding theory of reading and writing. Routledge, 2013.

[25] V. A. Thompson and A. Paivio, “Memory for pictures and sounds: Independence of auditory and visual codes.” Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, vol. 48, no. 3, 1994, p. 380.

[26] J. R. Binder, R. H. Desai, W. W. Graves, and L. L. Conant, “Where is the semantic system? a critical review and meta-analysis of 120 functional neuroimaging studies,” Cerebral cortex, vol. 19, no. 12, 2009, pp. 2767–2796.

[27] J. Wang, J. A. Conder, D. N. Blitzer, and S. V. Shinkareva, “Neural representation of abstract and concrete concepts: A meta-analysis of neuroimaging studies,” Human Brain Mapping, vol. 31, no. 10, 2010, pp. 1459–1468.

[28] R. E. Mayer, “Cognitive theory of multimedia learning,” The Cambridge handbook of multimedia learning, vol. 41, 2005, pp. 31–48.

[29] R. Mayer, “Share this page,” Journal of Computer Assisted Learning, vol. 33, no. 5, 2017.

[30] D. Mutlu-Bayraktar, V. Cosgun, and T. Altan, “Cognitive load in multimedia learning environments: A systematic review,” Computers & Education, vol. 141, 2019, p. 103618.

[31] R. J. Sternberg, Metaphors of mind: Conceptions of the nature of intelligence. Cambridge University Press, 1990.

[32] M. Kejriwal, Q. Wang, H. Li, and L. Wang, “An empirical study of emoji usage on twitter in linguistic and national contexts,” Online Social Networks and Media, vol. 24, 2021, p. 100149.

[33] U. Pavalanathan and J. Eisenstein, “Emoticons vs. emojis on twitter: A causal inference approach,” arXiv preprint arXiv:1510.08480, 2015.

[34] B. Stiegler. Philosophising by accident: Interviews with elie during (english edition). [retrieved: 4, 2017]

[35] M. C. Corballis, From Hand to Mouth: The Origins of Language. Princeton University Press, 9 2003.

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

[37] D. C. Dennett, From Bacteria to Bach and Back: The Evolution of Minds. W W Norton & Co Inc, 2 2018.