Addressing the Symbol Grounding Problem in VR

Muneo Kitajima [1] (Presenter) Makoto Toyota [2] Katsuko T. Nakahira [1]

[1]Nagaoka University of Technology, JAPAN
mkitajima@kjs.nagaokaut.ac.jp

[2]T-Method, JAPAN

A Short Resume of the Presenter

Muneo Kitajima is currently a professor emeritus at Nagaoka University of Technology. His recently published book "Memory and Action Selection in Human-Machine Interaction" (2016) proposes a unified theory of action selection and development by integrating PDP, Two Minds, and layered structure of human action. The theory provides a comprehensive view of how our brain functionally works in our daily life. His current interest is to understand the implications of the theory to development of skill of adaptive problem solving, the important skill for survival.

Project HP: http://oberon.nagaokaut.ac.jp/ktjm/organic-self-consistent-field-theory/index.html

HP: http://oberon.nagaokaut.ac.jp/ktjm/index.html

email: mailto:mkitajima@kjs.nagaokaut.ac.jp

Abstract

- A Virtual Reality (VR) environment presents objects that the user perceives and interacts with. It then transitions to the next state, reflecting the content of the interaction that has occurred as a result of the user's perception of the objects.
- For the interaction between the VR environment and the user to continue seamlessly, the meanings assigned to the objects in the VR environment by the creator of VR applications and the meanings held by the user experiencing them must be consistent.
- In this study, we propose a method to realize seamless interaction between VR environments and users by considering the objects presented by VR environments as symbols and capturing the relationship between the meanings they contain and the meanings held by the users who experience them through the symbol grounding problem, which is regarded as a challenging issue in the field of artificial intelligence.
- Based on the Mode Human Processor with Realtime Constraints (MHP/RT)[a], we focus on the fact that the content of human action selections is based on memes that are handed down from generation to generation and should provide a basis for his/her understanding of the situation of the surrounding world, and suggest that the symbol grounding problem can be solved by observing and identifying memes.

[a] MHP/RT is a cognitive architecture that can deal with action selections in everyday environments

PCM process

- Humans acquire information about the world through the five senses, and execute appropriate actions for the situation at hand with the Perceptual, Cognitive, and Motor (PCM) processes.
- Perceived information is encoded and represented as symbols, which makes it possible to think in the cognitive process. In the thinking process, memory is used to successively transform the symbols into new representations. Part of the result of thinking gives a representation of a sequence of actions performed in the motor process.

Physical Symbol System

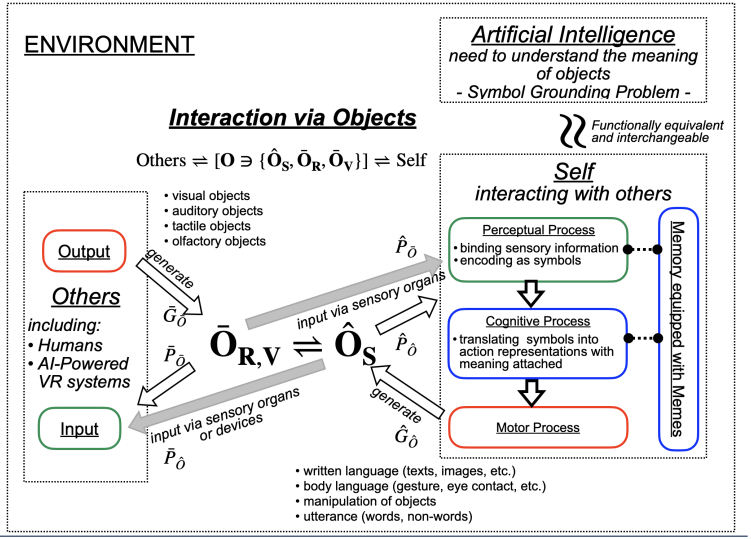
- Based on the idea that human intelligence can be captured by thought processes that manipulate symbols, Newell proposed the Physical Symbol System as a theory of human intelligence [1], providing the foundation for Soar [2, 3].
- When we try to artificially realize such interactions that humans perform in the real world using a physical symbol system, the system must have the ability to link the symbolic representations to the references in the real world and to acquire meaningful understanding from interactions with the environment.

- The realization of this capability is a fundamental challenge in Artificial Intelligence (AI) research and is referred to as the Symbol Grounding Problem (SGP) [4].
  - It concerns the ability of a machine to connect its symbolic representations to real-world references and acquire meaningful understanding from its interaction with its environment.
  - Addressing the SGP is crucial to building machines that can perceive, reason, and act like humans.
- One of the environments in which humans interact is a Virtual Reality (VR) environment.
  - The meaning that the user gives to the objects generated by the VR applications determines how the user interacts with the objects.
  - The VR applications can achieve a seamless interaction by having the ability to appropriately handle the meanings given by the user to the objects. *Here, it can be seen that the SGP is not unrelated to the realization of seamless VR environments.*

Three Types of Objects:

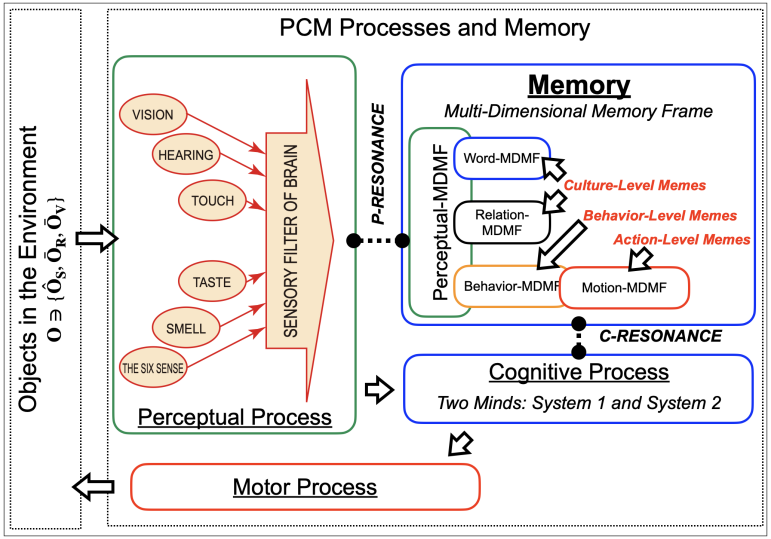
- **Self-Generated Objects $\hat{\mathbf{O}}_{\mathbf{S}}$:**
  Real objects that are directly generated by self, e.g., own speech, utterance, written text, drawings, gestures, modeling, etc.
- **Other-Generated Objects $\bar{\mathbf{O}}_{\mathbf{R}}$:**
  Real objects generated by other human beings with whom self is directly interacting.
- **Virtual Objects $\bar{\mathbf{O}}_{\mathbf{V}}$:**
  Objects generated by a machine or other human beings with whom self is not directly interacting.

Collectively represented as $\mathbf{O} \ni \{\hat{\mathbf{O}}_{\mathbf{S}}, \bar{\mathbf{O}}_{\mathbf{R}}, \bar{\mathbf{O}}_{\mathbf{V}}\}$, placed in the center of the figure.



Self, Others, and Artificial Intelligence:

- **Self:** Located on the right side of $\mathbf{O}$. Self operates the PCM processes to generate the objects $\hat{\mathbf{O}}_{\mathbf{S}}$ ($\hat{G}_{\hat{O}}$). Self also perceives them ($\hat{P}_{\hat{O}}$). The objects self perceives include the objects $\bar{\mathbf{O}}_{\mathbf{R,V}}$ generated by others ($\hat{P}_{\bar{O}}$).
- **Others:** Located on the left side of $\mathbf{O}$. Others generate the objects $\bar{\mathbf{O}}_{\mathbf{R,V}}$ ($\bar{G}_{\bar{O}}$). If the other is a human being, it runs the PCM process, which is equivalent to the one self runs, to generate the objects $\bar{\mathbf{O}}_{\mathbf{R}}$ ($\bar{G}_{\bar{O}_R}$). Machines run their own generative mechanisms to produce the objects $\bar{\mathbf{O}}_{\mathbf{V}}$ as output ($\bar{G}_{\bar{O}_V}$). The input to the others are the objects $\bar{\mathbf{O}}_{\mathbf{R,V}}$ generated by the others themselves ($\bar{P}_{\bar{O}}$), or the objects $\hat{\mathbf{O}}_{\mathbf{S}}$ generated by self ($\bar{P}_{\hat{O}}$).
- **Artificial Intelligence (AI):** Placed above Self. AI is functionally equivalent to Self and can be replaced.



PCM Processes and Memory:

- *PCM Processes and Memory* shown above is based on the Model Human Processor with Realtime Constraints (MHP/RT) [5, 6], 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.
- PCM processes use memory, modeled as Multi-dimensional Memory Frame (MDMF), consisting of Perceptual-, Behavior-, Motor-, Relation-, and Word-Multi-dimensional Memory Frame, abbreviated as P-, B-, M-, R-, and W-MDMF, respectively.

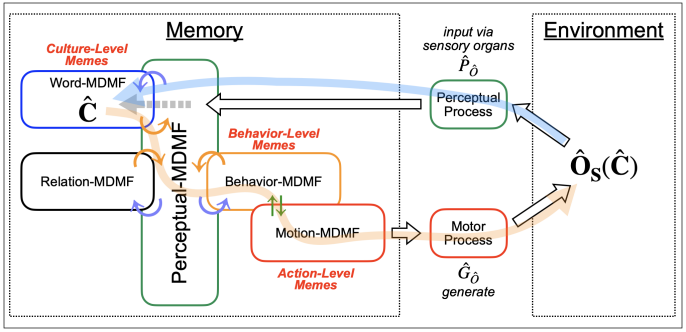
Features of MHP/RT and MDMF:

- P-MDMF overlaps with B-, R-, and W-MDMF; this unique configuration supports spreading activation from P- to M-MDMF, which connects perception with bodily movements.
- Perceptual information *resonates* with information in MDMF, called P-Resonance.
- In the cognitive process, conscious process utilizes W- and R-MDMF via C-Resonance, and unconscious process utilizes B- and M-MDMF via C-Resonance.
- The motor sequences are expressed according to M-MDMF.

Memory Organized as Memes: The basis of behavior is *imitation*; Behaviors that can be imitated across generations are preserved as sustainable behaviors. In this way, we can organize MDMF, which is used by the PCM processes and updated by their execution, in terms of *memes* that can be inherited across generations [7], in the following three levels of memes:

- *Action-Level Memes (A-memes)* represent bodily actions stored in M-MDMF.
- *Behavior-Level Memes (B-memes)* represent behaviors in the environment stored in B-MDMF.
- *Culture-Level Memes (C-memes)* represent culture stored in the R- and W-MDMF.

Interaction between Self and Self-Generated Object $\hat{O}_S$



This diagram shows the *Generate-and-Recognize* process and the situation where symbol grounding has been achieved within self, i.e., ***the SGP-solved situation***.

- *Generate*: Self generates an object $\hat{O}_S(\hat{C})$ that represents the concept $\hat{C}$ following the orange arrow through the motor process $\hat{G}_{\hat{O}(\hat{C})}$
- *Recognize*: Self recognizes the generated object $\hat{O}_S(\hat{C})$ following the blue arrow to reach the originated concept $\hat{C}$ through the perceptual process $\hat{P}_{\hat{O}(\hat{C})}$.

Generation Path [G-SS] (Orange Arrow):

$$\hat{C} \Rightarrow \left[\frac{\text{W-, R-, B-MDMF}}{\text{P-MDMF}} \rightleftharpoons \text{M-MDMF} \right] \Rightarrow \hat{O}_S(\hat{C})$$

Meme-Mapping from $\hat{C}$ to $\hat{O}_S(\hat{C})$:
The generation process [G-SS] can be rearranged from the perspective of memes, i.e., memories of contents, as follows:

$$\hat{C} \Rightarrow \left[\frac{\text{C-memes} \rightleftharpoons \text{B-memes}}{\text{P-MDMF}} \rightleftharpoons \text{A-memes} \right]_{\text{M-S}} \Rightarrow \hat{O}_S(\hat{C})$$

* M-S stands for Meme of Self

Recognition Path [R-SS] (Blue Arrow):

$$\hat{O}_S(\hat{C}) \Rightarrow \left[\frac{\text{P-MDFM}}{\text{W-MDMF}} \right] \Rightarrow \hat{C}$$

In [R-SS], if the concept $\hat{C}$ is strongly activated, then $\hat{O}_S(\hat{C})$ correctly realizes $\hat{C}$ in the real world. In this case, [G-SS] and [R-SS] are connected and closed, and the symbol $\hat{C}$ and $\hat{O}_S(\hat{C})$ are cognitively replaceable, which is represented by $\hat{C} \equiv \hat{O}_S(\hat{C})$. ***This state can be regarded as a state in which symbol grounding has been achieved within self***.

- The situation shown by slide 4 is solved by the fact that an activation pattern equivalent to the activation pattern of memes in MDMF that occurs in the process of generating self's $\hat{O}_S$ also occurs in AI.
- ***By ensuring that the meme-mappings occur within AI and self are equivalent, i.e.,*** $[\cdots]_{M\text{-}AI} \equiv [\cdots]_{M\text{-}S}$***, AI can be a substitute for self.***
- This is summarized in the flow shown in the right column, where Perceptual-Information-Encoding-in-AI (P-MDMF-AI) is the substitute for P-MDMF of human to perform A/D transformation to input the real world data to the AI system.
- Since the memes are knowledge passed down from generation to generation, it is considered possible to represent them by symbols.
- P-MDMF-AI can also be represented in symbols by encoding environmental information by sensors that perform the same function as sensory organs.
- ***The symbol grounding problem in AI is thought to be solved by elucidating memes.***

Symbol Grounding of Concept $\hat{C}$ in AI

1. The symbol $\hat{C}$, which is common to self's, in C-memes activates P-MDMF-AI as well as the associated C-memes in AI.
2. B-memes are activated via the activated portion of P-MDMF-AI.
3. The part of A-memes that overlap the activated B-memes is activated.

The steps 1, 2, and 3 constitute the meme-mapping of AI.

4. What is expressed by the activated A-memes is implemented in the real world via appropriate actuators.
5. Upon input of the object $\hat{O}_S(\hat{C})$ in AI, activation spreads in P-MDMF-AI, followed by the activation of the symbol $\hat{C}$ in C-memes.
6. The part in P-MDMF-AI that corresponds to $\hat{O}_S(\hat{C})$ and the symbol $\hat{C}$ integrate the C-, B-, and A-memes activated in the steps 1, 2, and 3 to form an integrated association. At this point, $\hat{C} \equiv \hat{O}_S(\hat{C})$ is established by AI by means of $[\cdots]_{M\text{-}AI}$, in other words, **the symbol $\hat{C}$ both the AI and the self commonly recognize has been grounded, secured by the relationship** $[\cdots]_{M\text{-}AI} \equiv [\cdots]_{M\text{-}S}$.

Interaction between Self and Other-Human-Generated Object $\bar{O}_R$

Generation Path [G-OO]:

[G-OO] ≡ [**G**enerate - symbol in **O**ther via memory of **O**ther]

$$\bar{C} \Rightarrow \left[\cdots \right]_{\text{MDMF of Other}} \Rightarrow \bar{O}_R(\bar{C})$$

Meme-Mapping from $\bar{C}$ to $\bar{O}_R(\bar{C})$:

$$\bar{C} \Rightarrow \left[\cdots \right]_{\text{M-O (Meme of Other)}} \Rightarrow \bar{O}_R(\bar{C})$$

Recognition Path [R-OS]:

[R-OS] ≡ [**R**ecognize object - generated by **O**ther using memory of **S**elf]

$$\bar{O}_R(\bar{C}) \Rightarrow \left[\frac{\text{P-MDFM}}{\text{W-MDMF}} \right]_{\text{MDMF of Self}} \Rightarrow \hat{C}$$

- If $\bar{C}$ ($\equiv \bar{O}_R(\bar{C})$) ≡ $\hat{C}$ holds, the symbol is transmitted through the object expressed by the other.
- Consider the communication via words, where the self and the other look at a sequence of words $\tilde{C}$.
  - *The self and the other perform symbol grounding according to their respective generation paths;* $[\cdots]_{M\text{-}S}$ *and* $[\cdots]_{M\text{-}O}$ *are included in each symbol grounding process.*
  - *If the self and the other have grown up in the same environment, which is the necessary condition for them to have a common set of memes, then the relation* $[\cdots]_{M\text{-}S} \equiv [\cdots]_{M\text{-}O}$ *would hold, and* ***the shared symbols have the same meaning****.*
  - *However, in the case of* $[\cdots]_{M\text{-}S} \neq [\cdots]_{M\text{-}O}$*,* ***the meaning of all visually shared symbols*** $\tilde{C}$ ***may not be shared****.*

Interaction between Self and VR-Generated Object $\bar{O}_V$

Generation Path [G-VV]:

[G-VV] ≡ [**G**enerate - symbol in **V**R via the mechanism in **V**R]

1. In a VR environment, the system takes in the information emitted by human users and then determines the response to it.
2. The input is represented as a symbol $\tilde{C}$.
3. Within the system, after setting a symbol $\bar{\bar{C}}$ to be transmitted in the next cycle of interaction, the symbol-object transformation is performed and the object $\bar{\bar{O}}_V(\bar{\bar{C}})$ is output to the user.

$$\bar{\bar{C}} \Rightarrow \left[\cdots \right]_{\text{Mechanism in VR}} \Rightarrow \bar{\bar{O}}_V(\bar{\bar{C}})$$

Recognition Path [R-VS]:

[R-VS] ≡ [**R**ecognize object - generated by **V**R using memory of **S**elf]

$$\bar{\bar{O}}_V(\bar{\bar{C}}) \Rightarrow \left[\frac{\text{P-MDFM}}{\text{W-MDMF}} \right]_{\text{MDMF of Self}} \Rightarrow \hat{C}'$$

$\hat{C}'$ represents the recognized symbol.

Generation Path [G-SS]:

[G-SS] is for generating image for the recognized symbol $\hat{C}'$.

- Upon recognition, the user (Self) activates his/her MDMF along the generation path [G-SS] for this symbol $\hat{C}'$ and obtains the corresponding object $\hat{O}_S(\hat{C}')$ in mind.
- ***If the relation*** $\hat{O}_{S(\hat{C}')} \equiv (or \approx) \bar{\bar{O}}_{V(\bar{\bar{C}})}$ ***holds, the interaction will proceed smoothly. If not, it will fail.***

Only if the object generation path [G-VV] *in the system is executed according to the user's meme mapping* $[\cdots]_{\text{M-S}}$*, i.e., if the meme mapping in the system is based on* $[\cdots]_{\text{M-AI}}$*, it is possible to proceed with an interaction that guarantees symbol grounding. Since memes are knowledge that are passed down from generation to generation, they can be represented by symbols. Explanations of memes and a method for externalizing memes are provided.*

Getting Memes into the Brain

- ***Action-Level Memes:*** During the period from birth to two to three years of age, humans generate a large number of synapses that connect neural circuits in the brain and take in as much information as possible. The rate of synapse generation is then reduced, and the distribution of information up to that point is used to determine the basic characteristics of the sensory organs. At the same time, by initiating body movements and imitating the movements of the people around them, they acquire body movements that have been formed empirically and accumulatively as individual ecology. This is formed through life's skillful method of adjusting the growth of muscles and other parts of the body to external constraints. At the same time, information from the sensory organs is linked to bodily actions. The most important bodily functions formed at this stage are the voice and hand functions.
- ***Behavior-Level Memes:*** Later, the voice paves the way to speech; the hands pave the way to tool use. Through continuous imitation, humans learn to use the words and tools of those around them. At this time, humans acquire a new hierarchy of actions by organizing and summarizing the fact that a particular collection of sounds evokes a particular response, and that the feel of a hand experienced through tactile sensation and the movement of a tool perceived visually are captured as an unified whole via the tool. This is made possible by linking the A-memes formed on the brain circuit according to the situation in which they are used, and making them available as a coherent whole.
- ***Culture-Level Memes:*** Furthermore, words pave the way to language, and tools pave the way to the use of more complex machines. At this stage, humans learn to act as members of the culture and civilization of the group to which they belong, not only through imitation but also through the experience of autonomous activities as members of the group. At this time, the B-level memes is extended to be used in a complex manner, and culture-specific behavior patterns are formed.

Mapping Memes into Information Systems

- The mechanism by which the memes inherit information is analogous to an information system.
  - *A-memes serve as the operating system that defines general patterns of spatial-temporal behavioral functions.*
  - *B-memes represent middleware 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.*
- By viewing memes as information systems, it is possible to represent human activities in various situations with symbols by observing them.
- Meme extraction has been attempted in some studies;
  - *The extraction of memes by observing the behavior of people trying to reach their destinations while acquiring information from information displays at railroad stations [8];*
  - *The inheritance of skills by ceramic artists through the acquisition of memes [9];*
  - *The memes used by skilled piano players during practice for a concert [10].*
- It is expected that memes in interaction in VR environments can also be elucidated based on these methods with appropriate modifications.

References I

[1] Allen Newell. **Physical Symbol Systems.** *Cognitive Science*, 4(2):135–183, 1980.

[2] Allen Newell. ***Unified Theories of Cognition (The William James Lectures, 1987).*** Harvard University Press, Cambridge, MA, 1990.

[3] Johnson E. Laird, Allen Newell, and Paul S. Rosenbloom. **Soar: An architecture for general intelligence.** *Artificial Intelligence*, 33:1–64, 1987.

[4] The symbol grounding problem. ***Physica D: Nonlinear Phenomena*, 42(1):335–346, 1990.** [retrieved: 2, 2025].

[5] Muneo Kitajima. ***Memory and Action Selection in Human-Machine Interaction.*** Wiley-ISTE, 2016.

[6] Muneo Kitajima and Makoto 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*, 5:82–93, 2013.

[7] Makoto Toyota, Muneo Kitajima, and Hideaki Shimada. **Structured Meme Theory: How is informational inheritance maintained?** In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, page 2288, 2008.

[8] Muneo Kitajima and Makoto Toyota. **Simulating navigation behaviour based on the architecture model Model Human Processor with Real-Time Constraints (MHP/RT).** *Behaviour & Information Technology*, 31(1):41–58, 2012.

[9] Muneo Kitajima, Makoto Toyota, and Jérôme Dinet. **Art and Brain with Kazuo Takiguchi.** In *COGNITIVE 2023 : The Fifteenth International Conference on Advanced Cognitive Technologies and Applications*, pages 1–10, 2023.

[10] Katsuko T. Nakahira, Muneo Kitajima, and Makoto 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*, 16:164–177, 2024.