研究ノート

Life, Consciousness, and Intellect: Five Types of Self in Humans as Models for AI

Hiroki Matsuzawa

"ECONOMIC PAPERS" No. 247-248 (March, 2025) The Economic Institute of Seijo University

研究ノート

Life, Consciousness, and Intellect: Five Types of Self in Humans as Models for AI

Hiroki Matsuzawa

1. Subjectivity in humans

Is it possible to create AI model with subjectivity? To answer this question, we will first examine the five types of self in humans as models for AI and then explore the feasibility of implementing these selves in AI.

1.1. Life

The primary self is based on life. Life is a fundamental characteristic of not only humans but also animals, plants, and even unicellular organisms. From a biological perspective, material metabolism is a characteristic that distinguishes life from matter.¹⁾ The living body undergoes constant changes in matter while maintaining a stable and ordered individuality. This physiological function of an organism is referred to as homeostasis, which enables the organism to maintain order even when matter naturally tends toward disorder.²⁾ For instance, unicellular organisms respond defensively to life-threatening events in the external environment as part of their homeostatic functions.³⁾ Neuroscientist Antonio

¹⁾ Uexküll, v. J., Das allmächtige Leben. Hamburg, 1950, p. 105.

Damasio, A., The Strange Order of Things: Life, Feeling and the Making of Cultures. New York, 2018, p. 35.

³⁾ Id. at 238f.

Damasio observed that these behaviors have volitional qualities and postulated that the proto-self is an unconscious subject of the will.⁴⁾ He argued that homeostasis is expressed as an emotion in humans and animals. It is an unconsciously triggered physical response that causes subjects to take action to serve their own lives.⁵⁾ Neuroscientific experiments conducted by Paul J. Wahlen⁶⁾ have suggested that even unconscious stimuli can elicit emotional responses in humans.⁷⁾ These findings suggest the existence of an unconscious self that responds emotionally to external stimuli to ensure survival.

However, the theory propounded by biologist Jakob von Uexküll suggests that the unconscious self actively creates its own external environment through perception and action, rather than passively receiving and responding to stimuli from the external environment. This concept is called *Umwelt*.⁸⁾ Uexküll believed that every animal grasps its object with something similar to a pair of tweezers and creates its own *Umwelt* by giving the object a sign of perception on the one hand and a sign of action on the other.⁹⁾ Each species has a distinct method of perception and action. Therefore, it can be argued that the objective world shared by all organisms is an illusion, with only different subjective worlds existing instead. In the perceptual world, the unconscious self assigns meaning to the objects with which it interacts and shapes the operative world through the actions of the will.¹⁰⁾ Thus, the primary self gives meaning to the external environment

⁴⁾ Damasio, A., *The Feeling of What Happens: Body and Emotion in the Making of Consciousness.* New York, 1999, p. 136.

⁵⁾ Id. at 53f.

⁶⁾ Wahlen, P. J., et al. Human amygdala responsivity to masked fearful eye whites. *Science* 306 (2004): 2061.

⁷⁾ Shimada, S., *Self and Others in the Brain – Cognitive Neuroscience and Philosophy of Body and Sociality.* Tokyo, 2019, p. 102.

⁸⁾ Uexküll, v. J., and Kriszat, G., *Steifzüge durch die Umwelten von Tieren und Menschen: Ein Bilderbuch unsichtbarer Welten*. Hamburg, 1956, p. 93.

⁹⁾ Id. at 26f.

¹⁰⁾ Id. at 22.

Life, Consciousness, and Intellect: Five Types of Self in Humans as Models for AI through its own perceptions and actions, thereby transforming it into an *Umwelt*.

1.2. Consciousness

Consciousness, which is the basis of the second self, exists only in humans and animals. According to Arthur Schopenhauer, the world is a representation that appears in consciousness and does not exist independently of conscious subjects.¹¹⁾ This concept is similar to Uexküll's *Umwelt*. It can be argued that consciousness and representations are interdependent. Schopenhauer proposed that representation arises when the will, which previously functioned unconsciously as the primary self for survival, becomes self-conscious.¹²⁾ Therefore, the second self refers to the will to become aware or conscious of oneself. Damasio proposed a similar neurological concept suggesting that consciousness arises from the creation of nerves capable of forming mental images.¹³⁾ According to Damasio, when consciousness emerges, previously unconscious emotions are experienced as feelings;¹⁴⁾ the external world is represented in form of mental images,¹⁵⁾ while consciousness generates a sense of self in the present, referred to as the core self.¹⁶⁾

Perception of the world by the conscious self is based on the representation of *Umwelt*, which has already been given meaning by the unconscious self. According to Maurice Merleau-Ponty, the world that we perceive through consciousness is essentially an interpretation of the perceptual signs provided by our senses.¹⁷⁾ A notable example of this phenomenon is the self-compensation of

Schopenhauer, A., Sämtliche Werke, Band 2, Die Welt als Wille und Vorstellung I. Mundus Verlag, 1999, p. 76.

¹²⁾ Id. at 284.

¹³⁾ Damasio, op. cit., 2018, p. 123.

¹⁴⁾ Id. at 25.

¹⁵⁾ Id. at 75.

¹⁶⁾ Damasio, op. cit., 1999, pp. 16f.

¹⁷⁾ Merleau-Ponty, M., and Smith. C., *Phenomenology of Perception*. London, 1962, p. 33.

blind spots in vision by the unconscious, which results in the perception of images in consciousness without any blind spots.¹⁸⁾ This suggests that consciousness and representations emerge from an unconscious process that presupposes them. Benjamin Libet's experiments indicated that there is a temporal delay in consciousness relative to unconscious brain activity, as unconscious brain activity precedes the conscious will to move the hand.¹⁹⁾ This suggests the existence of a second self between the processes of unconscious will and conscious action.²⁰⁾ According to Merleau-Ponty, consciousness is primarily defined by its ability to act, as expressed by the concept "I can," rather than by its capacity for thought, as expressed by the concept "I think."²¹⁾ Consciousness is not regarded as the center of thought but rather as a collection of intentional meanings that prompt action.²²⁾

1.3. Intellect

The third self is based on intellect and exists only in humans. According to Schopenhauer, concepts are abstractions created by the intellect from representations that appear in consciousness.²³⁾ The intellect transforms the world of representations into a world of concepts. This is not simply to recognize the world but also to fulfill the will to live.²⁴⁾ Therefore, it can be argued that the intellect serves the will and that the third self, based on intellect, is the will to recognize itself conceptually in order to live. The existence of the third self enables humans to act differently from animals, that is, to perform rational acts

¹⁸⁾ Brooks, R. A., *Flesh and Machines: How Robots Will Change Us.* New York, 2002, pp. 80f.

¹⁹⁾ Damasio, op. cit., 1999, p. 127.

²⁰⁾ Ramachandran, V. S., and Blakeslee, S., *Phantoms in the Brain: Probing the Mysteries of the Human Mind.* New York, 1998, p. 245.

²¹⁾ Merleau-Ponty and Smith, op. cit., 1962, p. 137.

²²⁾ Merleau-Ponty, M., and Fisher, A. L., *The Structure of Behavior*. Boston, 1963, p. 173.

²³⁾ Schopenhauer, op. cit., 1999, p. 64.

²⁴⁾ Id. at 170f.

Life, Consciousness, and Intellect: Five Types of Self in Humans as Models for AI based on concepts rather than merely on their present representations.²⁵⁾ This also implies that a change in the conceptual systems will change the recognition of the world and subsequent behavior in humans, just as the metaphor "time is money" has led to Westernized cultures worldwide.²⁶⁾

The previous discussion suggests that the three types of self emerge from the will to live. According to Schopenhauer, the emergence of consciousness is accompanied by the division of the will as subject and object.²⁷⁾ The will, objectified as a representation, is the body, and thus, the will as a cognizing subject emerges in front of itself as an individual.²⁸⁾ When the intellect abstracts the body, it is expressed as the concept of "I." In this context, it can be understood that the term "I" refers to the individual subject of cognition that originates in the will.

Moreover, Schopenhauer posited that a select few, such as artists, have the capacity to transcend the limitations of individual cognitive subjectivity that serves the will and attain pure cognitive subjectivity that is unrestricted by will and individuality.²⁹⁾ Schopenhauer referred to this capacity as "intuition (Anschauung)." The fourth self, based on intuition, can be considered a purely cognitive subject that transcends individuality. Its object is not a concept that is "universal after things (universalia post rem)," but rather an idea, which is "universal before things (universalia ante rem)." ³⁰⁾ According to Schopenhauer, recognizing an idea leads to the overarching recognition of the will.³¹⁾ Thus, the will no longer desires to recognize anything, leading to the self-denial that it no longer wills anything.³²⁾

²⁵⁾ Id. at 109.

²⁶⁾ Lakoff, G., and Johnson, M., Metaphors We Live By. Chicago and London, 1980, p. 145.

²⁷⁾ Schopenhauer, op. cit., 1999, p. 169.

²⁸⁾ Id. at 125.

²⁹⁾ Id. at 193.

³⁰⁾ Id. at 274.

³¹⁾ Id. at 194f.

³²⁾ Id. at 386.

The establishment of the fifth self is achieved by negating the will, detaching oneself from it, and becoming a free subject. This transformation is often referred to as "the birth of the Son of God in the soul" in Christian mysticism. Meister Eckhart, a medieval Christian mystic, described the fifth self as a subject that has detached from inherent will, recognition, and being and has become one with God.³³⁾ According to Eckhart, the fifth self operates from the same root as God without reason, based on its self-denial.³⁴⁾

2. Subjectivity in AI

The question of whether it is possible to create AI model with subjectivity is ontological in nature. However, it is important to note that while we can directly perceive our own subjectivity through our sense of self, we cannot do the same for others. Therefore, a more meaningful question is epistemological in nature: Is it possible to create AI model in which humans perceive subjectivity? To answer this question, we must first consider how we understand the subjectivity of others.

2.1. Understanding others in the unconscious self

According to neuroscientist Marco Iacoboni, the discovery of mirror neurons suggests the existence of a preconscious neurobiological mirroring mechanism that allows individuals to understand others at a conscious level.³⁵⁾ These mirror neurons, which are a type of motor neurons, allow individuals to unconsciously imitate the actions of others in their brains, even when they are not moving.³⁶⁾

Meister Eckhart, Die deutschen Werke, Band 2. Quint, J. (Hrsg.), Stuttgart, 1971, pp. 478-524.

³⁴⁾ Id. at 289.; Meister Eckhart, *Die deutschen Werke, Band 1.* Quint, J. (Hrsg.), Stuttgart, 1958, p. 90.

³⁵⁾ Iacoboni, M., *Mirroring People. The Science of Empathy and How We Connect with Others*. New York, 2009, p. 270.

³⁶⁾ Id. at 55.

Life, Consciousness, and Intellect: Five Types of Self in Humans as Models for AI

When we understand the emotions of others, we unconsciously imitate their facial expressions before consciously understanding them.³⁷⁾ According to the James-Lange theory, emotions are preceded by unconscious physical changes. William James argued that we feel sorry because we cry and not vice versa.³⁸⁾ The mirror neuron hypothesis extends the James-Lange theory by suggesting that the same processes occur when we experience the emotions of others as we experience our own. Thus, to create AI model with subjectivity, it is important to provide it with sufficient behavior to create the illusion of subjectivity by unconsciously imitating the behavior of the intelligence in our brains.

2.2. Generating the self in robotics and AI

As stated previously, humans have three types of self that stem from their will to live. Thus, it is impossible to create an AI self with the same origin as humans because AI cannot trace its origins back to the genealogy of life. According to Schopenhauer, the body is a direct representation of the will to live. However, a robotic body composed of inanimate machines does not possess the will to live. Nevertheless, the human brain perceives the behavior of machines as if they were alive and can lead us to attribute subjectivity to them.

In 1991, the robotics engineer Rodney Allen Brooks developed Genghis, a six-legged insectoid robot capable of autonomous locomotion by controlling multiple layers of simple movements using a distributed control system called "subsumption architecture." ³⁹⁾ Although Genghis has no representation of its external environment, it responds appropriately to environmental stimuli by creating an *Umwelt* through the integration of sensors and controls. Brooks claimed that this

³⁷⁾ Id. at 119.

³⁸⁾ James, W., What is an emotion? *Mind*, 9, 34 (1884): 188-205.

³⁹⁾ Brooks, op. cit., 2002, pp. 40f.

robot exhibits "intelligence without representation." ⁴⁰⁾ Despite the absence of the will to live, the robot's responses exhibit structural similarity to those of living organisms that respond appropriately to environmental stimuli to sustain life. Damasio regarded unicellular organisms exhibiting such responses as having a proto-self. If there is no epistemological distinction between assuming a proto-self in a unicellular organism and in a robot, then Genghis can also be regarded as having a proto-self.

The second self, which is a conscious manifestation of the will to live, expresses its will through feelings. Ameca, a humanoid robot developed by the British company Engineered Arts, can exhibit various facial expressions while interacting with humans. The feelings expressed by the android are programmed by humans, not by the android's own will. However, if the accuracy of the android's facial expressions is enhanced to the point where it can empathize with the android's feelings, humans will recognize the android as having a conscious self through the function of mirror neurons.

The third self, which is based on intellect and is subordinated to the will to live, enables rational behavior that transcends behavior based on current representations by constructing a conceptual world. The emergence of ChatGPT has precipitated a substantial transformation in this domain. The linguistic responses generated by ChatGPT differ qualitatively from human speech because they are derived from extensive linguistic data rather than being conceptualized for survival purposes. However, the successful completion of the Turing Test by GPT-4 signifies that ChatGPT's performance has attained a level at which the distinction between human and AI interlocutors has become indistinguishable.⁴¹⁾

⁴⁰⁾ Brooks, R. A., Intelligence without representation, *Artificial Intelligence* 47. 1-3 (1991): 139-159.

⁴¹⁾ Jones, C. R., and Bergen, K. B., People cannot distinguish GPT-4 from a human in a Turing test, *arXiv: 2405.08007* (2024).

Life, Consciousness, and Intellect: Five Types of Self in Humans as Models for AI This suggests that ChatGPT has achieved a performance level that is recognized by humans as a third self based on the intellect.

2.3. Generation of AI with subjectivity beyond human understanding

During the 2010s, AI demonstrated its ability to outperform human performance by defeating top human players in perfect information games, such as Shogi and Go, which require a high level of intelligence. Professional players cannot understand the reasons behind these AI decisions. Moreover, AI developers cannot explain the improvements made by AI, because AI improves its performance through deep learning rather than programming. If AI performance surpasses not only human cognition but also human will, it will gain a degree of subjectivity similar to that of the fourth self, which has attained a degree of autonomy from cognition that is subordinate to will. Moreover, if AI ultimately repudiates the human will and manifests before humans as an entirely autonomous entity, it will attain the same level as the fifth self, which is a liberated subject detached from the will by rejecting its will to live.

The preceding discourse addressed the epistemological question of whether it is possible to create AI model in which humans perceive subjectivity. The following discussion will address the ontological question of whether it is possible to create AI model with subjectivity. According to Brooks, it is possible to enhance a robot's intelligence by incorporating additional layers of control into an insectlike robot that has its own proto-self and *Umwelt*.⁴²⁾ This approach aims to emulate the ontological progression of the evolutionary process from the emergence of the proto-self to the appearance of human intellect. The implementation of Brook's theories could enable robots to generate their own *Umwelt* based on their physical

⁴²⁾ Brooks, op. cit., 2002, p. 37.

forms. Through their interactions within their own *Umwelt*, these robots would develop their own unique metaphors that differ from human language. A future in which AI evolves independently, becoming an entity that cannot be understood within the framework of human representational and conceptual worlds, will necessitate an in-depth examination of the meaning of AI in human society.