Artificial Pain: empathy, morality, and ethics as a developmental process of consciousness

Minoru Asada[0000−1111−2222−3333]

Osaka University, Suita, Osaka 565-0871, Japan
asada@ams.eng.osaka-u.ac.jp

Abstract. In this article, I propose a working hypothesis that the nervous system of pain sensation is a key component to shape robots' (artificial systems') conscious minds through the developmental process of empathy, morality, and ethics based on the MNS that promotes the emergence of concept of self (and others). First, the limitation of the current progress of AI focusing on deep learning is pointed out from a viewpoint of the emergence of consciousness. Next, the outline of ideological background on issues of mind in a broad sense is shown. Then, cognitive developmental robotics (CDR) is introduced with two important concepts; physical embodiment and social interaction both of which help to shape conscious minds. Following the working hypothesis, existing studies of CDR are briefly introduced and missing issues are indicated. Finally, an issue how robots (artificial systems) could be moral and legal agents is shown.

Keywords: Pain · Empathy · Morality · MNS.

1 Introduction

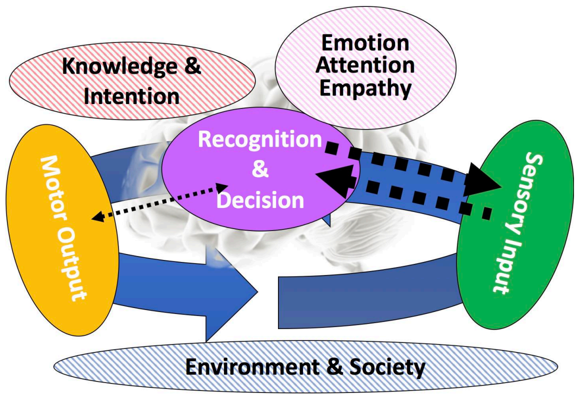

Fig. 1. The current status of deep learning

The rapid progress of observation and measurement technologies in neuroscience and physiology have revealed various kinds of brain activities, and the recent progress of AI technologies represented by deep learning (DL) methods is remarkable. Therefore, it appears no wonder that artificial consciousness can be realized soon. However, due to the fundamental limitations of the deep learning, it seems difficult. The main reason is that the current DL emphasizes the perception link between the sensory data and labels, lacking strong connection with the motor system, therefore, it does not seem to involve physical embodiment and social interaction both of which develop a rich loop including perception and action with attention, cognition, and prediction (Fig. 1). This is essential for consciousness research including unconsciousness.

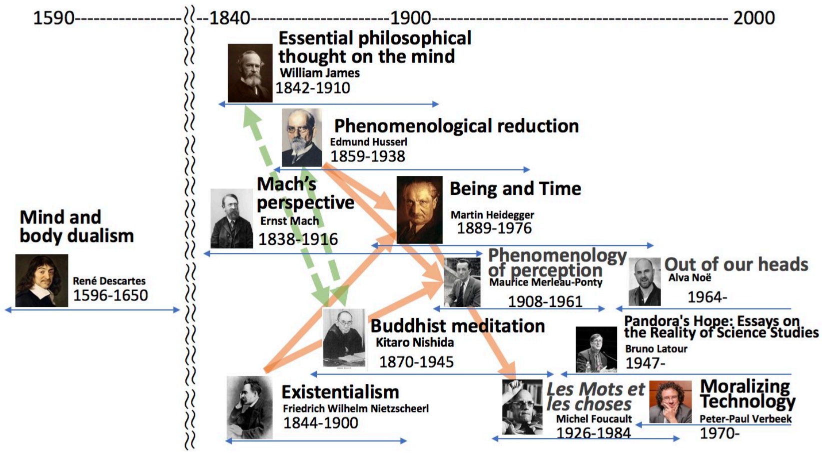

Fig. 2. Ideological background of relation between human being and things

Cognitive developmental robotics [2] has been advocating the importance of physical embodiment and social interaction, a big potential to overcome the above-mentioned limitation. The ideological background of constructive approach by CDR is well explained in his book by Jun Tani [9], chapter 3, featuring Husserl 's Phenomenology as follows (Fig. 2):

With regard to the relationship between the mind and body or things, it is Descartes who advocated mind and body dualism and laid the foundation of modern philosophy. It is Husserl who insisted on "New Cartesianism" that goes beyond Descartes to transcendental phenomenology and gave phenomenological consideration. He developed a way of thinking of subjectivity between subjective and objective intervals, and gave great influence to the next generations after him. He predicted that the analysis of nature is based on individual conscious experience. Heidegger and Merleau-Ponty have extended and evolved Husserl's phenomenological theory.

Heidega argues that "being-in-the-world" is born by dynamic interaction between the future possibilities of individual agents and their past possibilities without separating subjectivity and objectivity. He also pointed out the importance of some kind of social interaction that individuals can exist mutually under the prior understanding of how each individual interacts with purpose.

Merleau-Ponty argues that in addition to subjectivity and objectivity, the dimension of "physical embodiment" emerges, where the body of the same thickness is given to objects that are touched or viewed at the same time as the subject touching and seeing, and that the body could be a place where exchanges between two poles of subjective and objective is repeated. In other words, he pointed out the importance of the body as a media connecting an objective physical world and a subjective experience. As mentioned above, this is the basic concept of "physical embodiment" in the cognitive development robotics.

Based on these ideological backgrounds, CDR has done several studies, where computational models were proposed to reproduce the cognitive developmental processes by utilizing computer simulations and real robot experiments. Although CDR has not mentioned about consciousness explicitly, here we argue any possibility of artificial consciousness more explicitly by proposing a working hypothesis based on the nervous system of pain sensation. The story is as follows:

1. Embedding pain nervous system into robots for them to feel pain.
2. Through MNS (Mirror Neuron System) development, robots may feel pain in others.
3. That is, emotional contagion, emotional empathy, cognitive empathy, and sympathy/compassion develop inside robots.
4. Proto-morality emerge.
5. Robots could be agents who could be moral being, and at the same time, subjects to moral consideration.
6. Legal system for robots and AI will be considered.

The rest of the paper is organized as follows. First, the nervous system for pain sensation is briefly explained from a neuroscientific viewpoint. Next, a preliminary experiment of soft tactile sensor is shown as a potential of artificial nociceptor system. Then, we argue any possibility of artificial empathy, morality, and ethics in CDR by integrating existing studies and future issues. The above story can be regraded as a developmental process of artificial consciousness.

2 A nervous system for pain sensation

The perception of injurious stimuli, called nociception, or pain has its own nervous pathways different from mechanosensory pathways (see Chapter 10 in [8]). It transmits two kinds of information through the anterolateral system: the sensory discrimination of pain (location, intensity, and quality), and the affective and motivational responses to pain. The former terminates at somatosensory cortex (S1, S2) while the latter involves anterior cingulate and insular regions of cortex and the amygdala. The pain matrix consists of these four regions (Fig. 3). The analgesic effect arises from activation of descending pain-modulating pathways that project to the dorsal horn of the spinal cord from somatic sensory cortex through amhygdala and hypothalamus, then some parts of the midbrain, and regulate the transmission of information to higher centers. Such projections provide a balance of inhibitory (past view) and facilitatory influences that ultimately determines the efficacy of nociceptive transmission. In addition to the above projections, local interactions between mechanoreceptive afferents and neural circuits within the dorsal horn can modulate the transmission of nociceptive information to higher centers. This is gate theory of pain that explains the ability to reduce the sensation of sharp pain by activating low-threshold mechanoreceptors (kiss it and make it well).

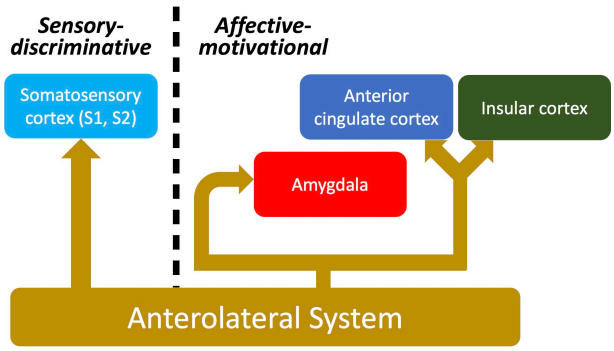

Fig. 3. Pain matrix (adopted from FIGURE 10.5 in [8])

3 From an artificial pain system to a moral being

3.1 Artificial pain

As a preliminary step of artificial pain nervous system, we developed a soft tactile sensor [4] consisting of four spiral inductors printed on a flexible printed circuit board (FPCB) and a disk-shaped magnetorheological elastomer (MRE; ferromagnetic marker) embedded in a cylindrical elastomer made of a silicon rubber. The inductances of the inductors are determined by the positional relationship between the ferromagnetic marker and each inductor because the marker contains iron particles with a high magnetic permeability. Therefore, the sensor can estimate applied tri-axis forces by monitoring the inductance changes caused by three-dimensional (3D) displacements of the marker. Fig. 4 shows the result of tactile sensation. The left two pictures show soft (rubbing) and hard (hammering) touches by a index finger and a hummer, respectively, and the right figure shows a time-course of three forces F_x, F_y, and F_z when the soft and hard touches were applied. As the figure shows, the sensor can discriminate soft and hard touches from the response waveform. The waveform for the hard touch is sharper than that of the soft one.

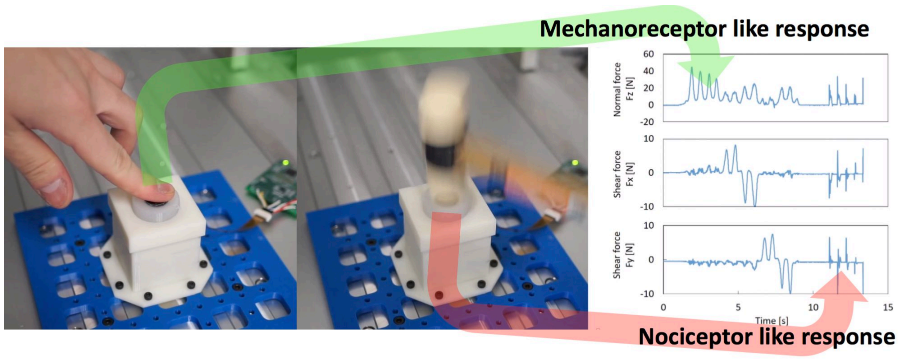

Fig. 4. A soft tactile sensor that discriminates soft and hard touches as a mechanoreceptor and a nociceptor, respectively

Based on the capability of discrimination of the tactile sensor, artificial nervous system for pain sensation can be embedded into the robot body and brain in parallel with normal mechanoreceptor pathway with a mechanism of pain regulation (the gate theory).

3.2 Artificial empathy

The feeling that this pain is shared is thought to be the source of empathy assuming that the mirror neuron system enables agents to percieve others' pain [1]. Actually, a survey article of artificial empathy [1] mentioned that many papers on empathy in neuroscience, cognitive science, and psychology dealt with pain as a research target. The right-bottom of Fig. 5 shows a conceptual model of empathy developmemt ([1]), and the rest of the figure indicate the related studies, which are briefly introduced in the followings.

Dynamic coupling between body and brain with neural oscillator networks Park et al. [7] showed that dynamic coupling between body and brain with neural oscillator networks generated two kinds of subnetwork structures apart from the anatomical network one; the former consists of many small

[1] https://www.nytimes.com/2006/01/10/science/cells-that-read-minds.html

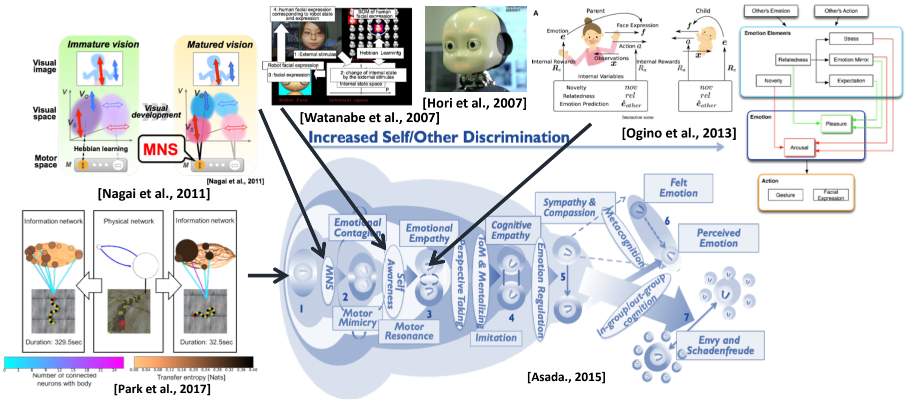

Fig. 5. A conceptual model of empathy development and related studies

subnetworks loosely connected each other and corresponds to a stable motion while the latter mainly consists of one big subnetowork strongly connected with sensory-motor nuerons and corresponds to an unstabel motion that connects the stable motions. The left-bottom of Fig. 5 indicates the summary of this work, and the following two points are essential for the main topic of the paper.

1. Two kinds of motions and network structures behind may correspond to very primitive lelvels of unconscious (stable motion) and conscious (unstable motion) states, respectively. More plausibly, stable motions could be attractors and unstable motions appear to transit between the attractors in the phase space.
2. The separation of two kinds subnetworks can be regarded as functional differentiation that is a basic mechanism for the emergence of new functions [11].

Emergence of mirror neuron system and emotion sharing Nagai et al. [5] proposed a computational model for the emegence of mirror neuron system based on the hypothesis that immature vision leads to self-other correspondence. At the beginnig, infants (robots) cannot discrimnate between self motion and others' ones due to their immature vision. Gradually, they become able to discrimnate owing to their visual development. However, early connections between action observation and action execution are left unchanged. As a result, the observation of both self-induced motion and other's motion evoke the motor system, that is, a function of the mirror neuron system.

Such mirroring could be expanded from action to emotions, that is, emotion sharing. Watanabe et al. [10] showed a computational model for emotion development based on a psychological finding, intuitive parenting. Starting from a very simple emotional space consisting of only two pleasure-displeasure states, an infant (robot) gradually differentiated its emotional space into richer one with happiness, surprise, anger, and so on, through the interactions with its caregiver. Hori et al. [3] proposed a unified model to estimate the emotional states of others and to generate emotional self-expressions by using a multimodal restricted Boltzmann machine (RBM). Ogino et al. [6] presented a motivation model of infant-caregiver interactions focusing on relatedness, one of the most important basic psychological needs that increases with experiences of emotion sharing. These three studies are positioned in Fig. 5.

Sharing painful situations induces sympathetic behavior In case of the emotion sharing of pain, the system needs to transmit two kinds of information, the sensory discrimination of pain (location, intensity, and quality), and the affective and motivational responses to pain as described in 2. The former information comes from the sensor system of the body while the second one comes from its own experiences of pain. The ideal story is as follows:

1. The information for sensory discrimination of pain (location, intensity, and quality) is transmitted to CNS from the sensory system.
2. If the above experience is the first time, the related information such as cause and/or reason is also transmitted with the information above.
3. Else, the memory of this experience is enhanced in the memory storage.
4. When the painful situation of others is observed, emotion sharing of pain happens, and also the memory of the similar experience is recalled.
5. Take actions to reduce the pain of others based on the recalled experience.

A robot could be a moral agent if it can generate such behavior successfully. At the same time, such a robot may have a right to receive moral behavior from others. Such a moral agency could be a solution to the first law of three laws of robotics [2]. That is, "A robot may not injure a human being or, through inaction, allow a human being to come to harm."

4 Discussion

To challenge the hard issue of consciousness, I attempted to represent it as a phenomena of the developmental process of artificial empathy for pain and moral behavior generation. The conceptual model for the former is given by [1] while the latter is now the story of fantasy. If a robot is regarded as a moral being who is capable to give moral behavior to others, is it deserving of receiving moral behavior from others? If so, can we agree that such robots have conscious minds? This is an issue of ethics towards robots, and also related to the legal system. Can we ask such robots a sort of responsibility for any accident they committed? If so, how? These issues arise when we introduce robots who are qualified as moral being with conscious minds into our society.

Before these issues, there are so many technical issues. Among them, the followings should be intensively addressed.

[2] https://en.wikipedia.org/wiki/Three_Laws_of_Robotics

1. Associate the sensory discrimination of pain with the affective and motivational responses to pain (the construction of pain matrix and memory dynamics).
2. Recall the experience when a painful situation of others is observed.
3. Generate appropriate behavior to reduce the pain.

Acknowledgments

This research was supported by JST Strategic Basic Research Programs (RISTEX), Research Area "Human-Information Technology Ecosystem," entitled "Legal Beings: Electronic personhoods of artificial intelligence and robots in NAJIMI society, based on a reconsideration of the concept of autonomy"(Oct. 2017 Sep. 2020)

References

1. Asada, M.: Towards artificial empathy. International Journal of Social Robotics **7**, 19–33 (2015)
2. Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., Ogino, M., Yoshida, C.: Cognitive developmental robotics: a survey. IEEE Transactions on Autonomous Mental Development **1**(1), 12–34 (2009)
3. Horii, T., Nagai, Y., Asada, M.: Imitation of human expressions based on emotion estimation by mental simulation. Paladyn, Journal of Behavioral Robotics **7**(1), 40–54 (2016)
4. Kawasetsu, T., Horii, T., Ishihara, H., Asada, M.: Flexible tri-axis tactile sensor using spiral inductor and magnetorheological elastomer. IEEE Sensors Journal **18**(4), 5834–5841 (2018)
5. Nagai, Y., Kawai, Y., Asada, M.: Emergence of mirror neuron system: Immature vision leads to self-other correspondence. In: IEEE International Conference on Development and Learning, and Epigenetic Robotics (ICDL-EpiRob 2011). pp. (CD–ROM) (2011)
6. Ogino, M., Nishikawa, A., Asada, M.: A motivation model for interaction between parent and child based on the need for relatedness. Frontiers in Psychology **4**(Article618), 324–334 (2013)
7. Park, J., Mori, H., Okuyama, Y., Asada, M.: Chaotic itinerancy within the coupled dynamics between a physical body and neural oscillator networks. PLOS ONE **12(8)**, 618–628 (2017)
8. Purves, D., Augustine, G.A., Fitzpatrick, D., Hall, W.C., LaMantia, A.S., McNamara, J.O., White, L.E. (eds.): Neuroscience, fifth edition. Sinauer Associates, Inc. (2012)
9. Tani, J.: Exploring Robotic Minds: Actions, Symbols, and Consciousness as Self-Organizing Dynamic Phenomena. Oxford University Press (2016)
10. Watanabe, A., Ogino, M., Asada, M.: Mapping facial expression to internal states based on intuitive parenting. Journal of Robotics and Mechatronics **19**(3), 315–323 (2007)
11. Yamaguti, Y., Tsuda, I.: Mathematical modeling for evolution of heterogeneous modules in the brain. Neural Networks **62**, 3–10 (2015)