## [Grant-in-Aid for Scientific Research (S)]

# Constructive Approach to the Emergence of Moral Sense Based on a Baby Body-Brain Model with Internal Organs

	Principal Investigator	The University of Tokyo, Graduate School of Information Science and Technology, Professor	
		KUNIYOSHI Yasuo	Researcher Number : 10333444
170	Project Information	Project Number : 25H00448	Project Period (FY) : 2025-2027
		Keywords : Embodiment, Cognitive Development, Neuroaesthetics, Somatic Markers	

#### Purpose and Background of the Research

#### • Outline of the Research

With AI rapidly advancing beyond human-level intelligence, the question of humanity and embedding moral values into AI has become critical. This research integrates cognitive developmental robotics, developmental cognitive science, and neuroaesthetics to explore how primitive moral values emerge from embodiment. A baby body-brain model will be enhanced with an organ model for emotional foundations, enabling spontaneous actions and learning in a virtual environment. Human experiments will examine the development of moral values, aesthetic sense, and empathy. By integrating these approaches, we aim to understand the mechanisms behind moral emergence and explore AI implementation, validated through robotics. Unlike traditional methods, an embodiment-based moral model is deeply embedded in cognitive functions, ensuring adaptability to unforeseen situations. Our models and methodologies will be shared with global academic communities to foster collaborative research.



• Developmental Construction (Hypothesis) Figure 1 (right) illustrates our hypothesis: the green pathway from the musculoskeletal system and movement merges with the reddish-brown pathway from the internal organs and metabolism, leading to moral sense. Developmental science, neuroaesthetics, and constructivists collaborate to elucidate each component and its connections.

### Expected Research Achievements

• Embodied Development Can moral "seeds" be embedded into AI's core principles, ensuring all behaviors and decisions adhere to them without loopholes? The key may lie in development based on embodiment. "Embodiment" is not just a physical interface for AI but a principle enabling intelligent behavior in all situations, including unforeseen ones. The musculoskeletal system, for example, structures movement and sensation by enforcing constraints rather than prescribing specific actions. Extending this idea, embodiment shapes the interactions between the musculoskeletal, internal organs, metabolic/neural systems with the environment and other agents, as well as between the internal components of the system. Through developmental construction, spontaneous sensorimotor learning fosters elementary interactions, which evolve into higher-order structures, giving rise to perception, behavior, cognition, and human-like traits. All emerging functions follow embodiment constraints, ensuring infinite inputoutput variations conform to specific principles. If based on human embodiment, AI should share a fundamental structure with humans. This forms the foundation for constructing AI with a robust and loophole-free moral sense, effective beyond training data.

• Early Development of Moral Sense, and Neuroaesthetics How does moral sense emerge through continuous development under embodiment? Recent developmental cognitive science studies show that even pre-linguistic infants prefer justice and exhibit punitive actions against immoral agents (Kanakogi+ 2017, 2022). This suggests that moral sense originates from sensorimotor learning before verbal instruction. Another clue linking morality and the body comes from neuroaesthetics. Classical philosophy considers truth, goodness, and beauty as three virtues humans should pursue [Plato, Symposium], with the ideal moral person embodying both goodness and beauty [Plato, Gorgias]. Neuroaesthetic studies show that both goodness and beauty activate the orbitofrontal cortex, suggesting a close relationship (Ishizu & Zeki, 2011; 2013; 2014; 2017). This brain region is part of the emotional system and interacts with visceral functions, implying a strong embodiment basis. Neuroscience has long recognized the impact of interoception and the autonomic nervous system on cognition and emotion. Recently, brain-gut interactions, particularly in stress-related irritable bowel syndrome, have gained attention. However, the precise mechanisms linking the brain and visceral activity remain only partially understood. We aim at revealing and modeling them with a focus on early development.

• Research Plan The ultimate goal is to reconstruct the entire process in Figure 1 (right) and model its spontaneous development from the fetal stage. This will identify key elements and generalize them into artificial system design principles. In this project, we focus on two developmental pathways: the reddish-brown route (interoception  $\rightarrow$ emotion  $\rightarrow$  aesthetic & moral judgment  $\rightarrow$  moral sense) and the green route (movement & proprioception  $\rightarrow$  self-other cognition  $\rightarrow$  action understanding  $\rightarrow$  moral sense). "Selfembodiment" (including visceral/metabolic and interoceptive senses) forms the foundation. From this, emotion & preference arise, passing through aesthetic & moral judgment, leading to moral sense. "Movement" and exteroceptive & somatosensory perception allow differentiation of the self from others, forming empathy, self-other recognition, and imitation, which lead to action & intention understanding. Both pathways integrate into moral sense. Experiments will first test adults, then infants, and finally robots for validation. A unified theory will be constructed, recreating Kanakogi et al.'s infant moral sense experiments (Kanakogi+ 2017). The final goal is to develop a model that exhibits infant-like moral responses, implement it in robots, and verify its effectiveness.

Homepage Address, etc.	https://www.isi.imi.i.u-tokyo.ac.jp/ https://baby-lab.hus.osaka-u.ac.jp/ https://wps.itc.kansai-u.ac.jp/pan-lab/ https://www.daikoku-lab.com/about
---------------------------	---