

Robot as Moral Agent: A Philosophical and Empirical Approach

Shoji Nagataki

Chukyo University

Masayoshi Shibata

Kanazawa University

Tatsuya Kashiwabata

Keio University

Takashi Hashimoto

Japan Advanced Institute of Science and Technology

Takeshi Konno

Kanazawa Institute of Technology

Hideki Ohira

Nagoya University

Toshihiko Miura

University of Tokyo

Shinichi Kubota

Kanazawa University

Abstract: What is necessary for robots to coexist with human beings? In order to do so, we suppose, robots must be moral agents. To be a moral agent is to bear its own responsibility which others cannot take for it. We will argue that such an irreplaceability consists in its having an inner world — one which others cannot directly experience, just as pleasure and pain. And personality of a moral agent, which is to be irreducible to a mere difference of traits or features of individuals, is firmly rooted in such an inner world.

We will support our theses by referring to our experiment in which humans and robots interact with each other doing a coordination task. This experiment will provide an empirical analysis of the human-robot relationship with regard to learning mechanism, moral judgement, and the ascription of the inner world.