"Involving Interface": An Extended Mind Theoretical Approach to Roboethics

Citation for published version:
Anderson, M, Ishiguro, H & Fukushi, T 2010, '"Involving Interface": An Extended Mind Theoretical Approach to Roboethics', *Accountability in Research: Policies and Quality Assurance*, vol. 17, no. 6, pp. 316-329. https://doi.org/10.1080/08989621.2010.524082

Digital Object Identifier (DOI):
10.1080/08989621.2010.524082

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Accountability in Research: Policies and Quality Assurance

"INVOLVING INTERFACE": AN EXTENDED MIND THEORETICAL APPROACH TO ROBOETHICS

Miranda Anderson (University of Edinburgh), Hiroshi Ishiguro (Department of Systems Innovation, Osaka University & ATR Intelligent Robotics and Communication Laboratories) & Tamami Fukushi (Research Institute for Science and Technology for Society, Japan Science and Technology Agency)

In 2008 the authors held "Involving Interface" a lively interdisciplinary event focusing on issues of biological, sociocultural and technological interfacing (see acknowledgments). Inspired by discussions at this event, in this paper we further discuss the value of input from neuroscience for developing robots and machine interfaces, and the value of philosophy, the humanities and the arts for identifying persistent links between human interfacing and broader ethical concerns. The importance of ongoing interdisciplinary debate and public communication on scientific and technical advances is also highlighted. Throughout the authors explore the implications of the extended mind hypothesis for notions of moral accountability and robotics.

Keywords: *extended mind; ethics; robotics; interface; interdisciplinary; moral accountability; mirror neurons; neuroscience; public communication.*

The purpose of "Involving Interface" was the wider dissemination and encouragement of exchanges between disciplines, especially in order to keep interactive ongoing debates about the history and evolution of biological, technological, and sociocultural interfacing, and to consider the wide-ranging ethical implications. The founding premise of this event is also the basis of this paper: the notion that sociocultural, technological and physical factors in combination *constrain and enable* human cognitive capacities, including notions of morality. Therefore an extended mind approach is suggested here as important to considerations of moral accountability, both in general and specifically in relation to robotics. An extended mind approach emphasizes the meshing of a mass of interactive factors in making up cognitive (and moral) agents and is linked to notions of hybridity and diversity. The authors believe that a foundational implication of an extended approach is that robotic entities should not just be confined to current or human-like forms of morality as a goal. Firstly, perceptions of morality contain elements that vary, as well as those that persist, over geographical and temporal spans. Secondly, given the increasingly diverse array of types and complexities of robots (as of life forms), and of their intended purposes and environments, different forms of moral code will for that reason need to be developed and implemented, whether based on type-specific models or on a general model which can then be accordingly adapted. Lastly, while in the distant future "roboethics" may need to address more urgently the legal issues of robots' rights and research guidelines on experimentation on robots, already this approach invites discussion of how robots and machine interfaces are poised to reflect and interact with the creating of ethical categories and concepts of accountability.

Diverse notions of accountability underlie the funding, performance, and presentation of research through its interfacing with public, institutional and commercial domains. Rather than focusing on particular disciplinary methodologies, this paper tackles broader issues of accountability in relation to changing definitions of human and biological nature as plastic and extendable into the world. These changing definitions make a case for more communication over research in different fields on the interconnected factors that make up living beings and, the authors suggest, implies the need for an interdisciplinary approach to the issue of roboethics. Interdisciplinary debates invite us to imagine the variety and complexity of roles robots could play in our world. Multiple perspectives, which can be illuminated both through theoretical arguments and through practical experimentation, are necessary in the processes of understanding our selves, our world and our robots. Artistic creativity also allows less restricted forms of thinking to emerge on these topics. For example, the artist Stelarc through his use and incorporation of prosthetic devices has done much to challenge traditional notions of body boundaries, establishing a concept of authenticity as relating not to coherency or individuality, but to the ability to collaborate and to connect (2005).

The remainder of this paper considers a few of the broader theoretical implications of an extended mind approach for ethical robotic systems, through taking into account the relation of neuroscience to robotics, and the contributions of literature, interdisciplinary relations and public communication.

An Extended Mind Approach

The hypothesis that the mind is extended into the world is becoming increasingly influential (Hutchins, 1995; Clark and Chalmers, 1998; Clark, 1997, 2003, 2008b). The authors presume that both the "mind" and "subject" are metaphorical concepts: they therefore extend more fluidly into the world than the biological brain or body structure, which contribute to their conceptual capacity. The mind cannot be literally described: even those who would reduce it to the biological brain are hampered by the current incompleteness of knowledge of its physical nature, and there is a manifest difference in scale between neural activity and the capacities of the mind. Therefore analysis of the mind should take into account not only the findings of current neuroscience but also the metaphors that shape and are shaped by the biological nature of the brain, language, and sociocultural and technological trends, both for the ways these metaphors are helpful and for the ways they may lead us astray. Similarly, the subject is not reducible to the biological organism, although its body is a participating factor in its formation. Scientific terms and cultural and literary history are already implicated in each other. This holds implications for how one conceives of ethics generally, and of roboethics specifically, inviting input from across the range of academic disciplines and creative art forms.

As well as forms of extendedness that relate just to cognitive processes, it is worth considering the more general use of other people, language, objects and robots as an extension of the subject, called here "extended subjectivity", and the various ways in which such extendedness can be a recursive means to self-knowledge via an "extended reflexivity" (Anderson, 2007a). Stelarc only makes visible by taking one step further the extent to which all human minds and subjects loop out into other subjects, the environment and technologies. Intimate forms of technological, sociocultural and linguistic interfacing are so habitual they are often invisible to us;

right now "our" words on the page are resounding in your mind, leaving the permeability of boundaries revealed by their fluidly intersubjective operation, between people, and intrasubjective operation, within a person. Experiencing the breakdown of a laptop or the loss of a friend quickly reminds us just how fundamentally distributed we are. Stephen Kosslyn defines as "social prosthetic systems" (SPSs) people that we "rely on to extend our reasoning abilities and to help us regulate and constructively employ our emotions"; he suggests that we have evolved these social systems for the same reason that Clark explains that we employ bodily and technological resources: because our brains are limited. (Kosslyn, 2005, 2006; Clark, 2003).

The biological body structure is also limited but the subject can experience non-biological resources as part of the body itself. One of the authors, Hiroshi Ishiguro, experiences an extended body image via his robotic interface Geminoid, which suggests that work being done on the haptic sense and on the teletransportation of self-perspective may find interesting ways of expanding its research through collaborations with roboticists (Ishiguro, 2005; Ehrrson, 2008; Berti and Frassinetti, 2000; Maravati and Iriki, 2004; Tsakiris, 2008). More generally, robots, and especially androids or humanoids, can act like mirrors: by figuring a representation of living forms, they are potentially revelatory of the working of our interior and social worlds. They are neither entirely subject nor entirely object, and this raises questions and tensions about our own liminality and heteronomy (Anderson, 2007b). Historical displacements onto women, other nationalities, the emotions, and the body as a means of purging and shoring up the leaky walls of human subjectivity now find in the cyborg a spectre that appears to threaten humans' constitutive supremacy by its reflection of our own inherent hybridity, which results in our fear and fascination over other liminal forms. Yet, transformability is part of our natural inheritance, descended together with technologies, from our evolutionary past. The pertinence of this debate is heightened by robotics, microprocessing, nanotechnological, biotechnological and genetic engineering advances (Gillet, 2007; Ross, 2006; Landecker, 2005). These advances result in our increasing ability to add to or manipulate both our biological and non-biological resources. New technologies require and create new categories and concepts, as well as transforming those we apply to ourselves and other life forms (Turkle, 2004).

Yet, most research in orthodox cognitive science remains "recognizably Cartesian in character": while Cartesian substance dualism has been widely rejected, Cartesian psychology continues to shape work in cognitive science in terms of a number of principles, including an "explanatory dualism" that posits "a divide between mind and the rest of nature" (Wheeler, 2005). The extended mind hypothesis holds potential consequences for all sectors of society: "This is a confrontation long overdue, and it is one with implications for our science, morals, education, law, and social policy" (Clark 2003). However, Clark does not appear to tackle the question of how his rejection of the traditional "executive self" in favour of a coalition "soft self" can be made responsible within society other than his implicit trust in the value of acknowledging this as the true make up of the human subject. In response to Alice Juarrero's questioning of "how responsible agency is to be fleshed out" (Juarrero 2004), once it is allowed that it is "tools all the way down", Clark suggests he as yet does "not have a good answer" (2004).

Yet, as Clark has argued in the past, ethical theory tends to concern itself with the individual in relation to society, reminding us that "moral reason involves crucial collaborative, interpersonal dimensions" (1996). In emphasizing that linguistic principles are part of the mechanism of moral reason rather than just imperfect mirrors of moral knowledge, Clark warns that: "Oversimplified connectionist models of moral cognition, by marginalizing the collaborative dimensions of moral action, likewise threaten to isolate the moral agent from her proper home, the moral community" (1996). Similarly highlighting the collaborative and interpersonal nature of morality the cultural theorist Judith Butler argues that humans' lack of a unified executive self is not a hindrance to, but the grounding for responsibility: "my own foreignness to myself is, paradoxically, the source of my ethical connection with others" (2001). Conversely, the individualist morality that Butler critiques, by cutting off the subject from the world, destroys the basis for its moral engagement with it (2005). Social responsibility arises from interrelationality rather than a unified self. Yet Butler's argument remains focused on sociocultural and linguistic structures, where it might be better argued that biological, sociocultural and technological factors all contribute to human ethical dispositions and practices. The construction of complex moral agents will need to fully take into account the distributed nature of human morality and will require functional equivalents of the material factors. More complex moral agents will benefit from humanlike appearance and movements, due to our tendency to assume that such agents have higher cognitive and moral capacities (Farah and Heberlein, 2007; Krach et al., 2008), but these may be created through different matter and means, and more generally while a body is required it is not necessarily a body "just like ours" (Clark 2008a).

In sum, the points discussed indicate that a hybrid approach aimed at achieving a reflective equilibrium that incorporates and integrates both top-down rules and bottom-up learning, developmental and evolutionary mechanisms will be necessary in artificial moral agents that aim at full and flexible moral agency with the built in capacity for forging ethical connections through other beings and robots. Meanwhile more basic moral agents will need to have at least what Wallach and Allen have called "functional morality", that is designers and users that consider the ethical values being implemented in and through robots, since ethical values are created in robots and designed systems whether or not the engineer explicitly constructs them; therefore at the bottom level awareness of this needs to be raised (2009).

From Neuroscience and Neuroethics to Robotics and Roboethics

> "What we make" and "what (we think) we are" coevolve together; emergence can operate as an ethical dynamic as well as a technological one. (Hayles, 2005).

A claim that the things made by us are intimately related to what we make of ourselves would appear to err by confusing the producer with the product. Humans, understood as conscious and rational agents, are in opposition to objects, understood as without reason, will or consciousness. In the liberal humanist model the role that objects play is as passive matter without our reasoning capacities; as puppets moved by our commands; and, at most, as the external result of internal mental cogitations. The most recent challenges to this model of the human subject as autonomous intellect have been prompted by discoveries in neuroscientific research which, in combination with theoretical developments in cognitive science, have revealed the subject to be fundamentally shaped by its interactions, tools and creations, as well as the shaper of them. It is increasingly being demonstrated that the neurobiological mechanisms which participate in our concepts of the mind and of the subject, rather than acting just as a limitation on hybridity and extendability, in fact have a plasticity and interrelationality that invites dynamic and intimate relationships with the world (Berti and Frassinetti, 2000; Clark, 2003; Gallese et al., 2004; Maravati and Iriki, 2004; Noë, 2004). The field of neuroscience over the last two decades has been instrumental in making apparent how much of human cognition takes place through loops out into the body, technologies, other subjects and the environment.

Although diversity of opinion and fertile metaphors abound about exactly how the brain works, the general idea followed by artificial architectures involves interlinked activations of a mass of parallel processing units that are distributed across the brain. Patterns of neural activity are generated in response to excitatory or inhibitory inputs, caused by the synapses' modulating effects, and this activity in turn modifies the synapses themselves. It has become accepted knowledge and a motivating force in neuroscientific research that experiences modify not only the activity but also the organization of neural circuitry: "One of the most important and fascinating properties of the mammalian brain is its plasticity; the capacity of the neural activity generated by an experience to modify neural circuit function and thereby modify subsequent thoughts, feelings and behaviour" (Citri and Malenka, 2008). Such evidence of neural plasticity suggests that the brain is poised to be shaped by, as well as the shaper of, technologies and the surrounding environment.

Another potential contribution of neuroscience to the study of ethics is to provide quantitative, more objective procedures that can aid in assessing the relation of brain functions associated with certain mental states and with "personhood" to ethical concerns. Personhood is a foundational concept in ethics yet neither psychological traits nor neuroscientific evidence is sufficient to define what it is in itself. We cannot explain through quantitative evidence why we define that a given agent has personhood, even although brain mechanisms cause us to intuitively assess and ascribe a given agent with personhood. Whilst attempts to use brain function alone may seem to potentially entail a reductionism of personhood, neuroscientific tests in fact have even more surprisingly provided evidence of the illusory nature of this concept (Farah and Heberlein, 2007). Due to this, neuroscience can also be a useful tool to explain the "uncanny valley" hypothesis, first described by the Japanese roboticist Masahiro Mori (1970), as it is non-conscious human brain functions that lead to the projection of "personhood" on androids, humanoid robots or equivalent agents. This leads to expectations of high cognitive functioning, with these entities' current failure to satisfactorily achieve this then leading to a sense of uncanniness in the human spectator (Krach et al., 2008). In any case, even although neuroscience can give us some measures to test whether the given object is recognized as a person, rather than leading to a definition of personhood per se, it instead arguably makes the case for the consideration of other measures of accountability.

Neil Levy in his introduction to the first issue of *Neuroethics* also discusses ways in which neuroscience is placing in question traditional assumptions about human rationality, autonomy and morality (2008; see also Fukushi et al., 2007). What constitutes "humanity" is constantly laid open to question by our tendency towards hybridity and relationality, and this tendency is contributed to by the input of neurobiological mechanisms. The research of Rizzolatti and his colleagues has led to the astonishing discovery of a brain system which further demonstrates that our cognitive processes are not exclusively centred within a container subject: "Mirror neurons represent the neural basis of a mechanism that creates a direct link between the sender of a message and its receiver" (Rizzolatti and Craighero, 2004; see also Rizzolatti and Sinigaglia, 2008). Meanwhile the work of Antonio Damasio and his colleagues has further established the role of body states and emotions in reasoning and social inference, and the role of non-conscious processes in conscious decision making (1994). Nor is it just humans who are capable of such fluidity concerning physical and identity structures: J. Scott Turner has shown in his ecological research that even basic life forms commonly use social and environmental offloading and he contends that "animal-built structures are properly considered organs of physiology" (2000).

Yet while mirror neurons provide evidence of a vehicle that biologically extends human subjects (based on the Rizzolatti's view), the mental state does not remain identical across first and third person boundaries without the intervening factor of shared first-person experience (Calvo-Merino et al., 2005). More general neuroscientific studies suggest that specific enactive subjective experience frames future experience. Agloti and his colleagues' study on action anticipation and motor resonance in trained basketball players demonstrated that "the fine-tuning of specific anticipatory 'resonance' mechanisms" endow elite athletes' brains "with the ability to predict others' actions ahead of their realization" (2008). Thus, while mirror neurons indicate the potential for considerable sharing of various types of experience across persons, this evidence suggests that there is also considerable particularity of subjective experience, and that enactive cognition plays a significant role in forming our cognitive repertoire and mirroring potential.

Our subjective experiences of the world are made up of a rich and dynamic mix of shared consistencies and particular divergences. Iacobini posits that the recent discovery of "super mirror neurons" which play a modulating role in mirror neuron activity by inhibiting overt copying, is one of the mechanisms which allow the distinction between self and other to emerge (Iacoboni, 2008). On similar lines, at "Involving Interface" Ikegami made the vital point that intimate interfacing involves attachment and detachment simultaneously, and in discussing the implications of current research on mirror neurons for Artificial General Intelligence emphasized their revealing the importance of social empathy, co-creativity and collaborative behaviour. These abilities are a vital part of the basis of ethical systems.

An architectural basis of affective states and processes, such as is being developed in the reactive mechanisms of Sloman and Chrisley's model CogAff, would be necessary in a sophisticated artificial moral agent (Sloman et al., 2005). So would sophisticated mechanisms for sharing across and distinguishing first and third person perspectives, some of the basics of which appear to be emerging through the work of Scasselatti and Breazeal, respectively working on Nico, a robot who can distinguish between his mirror-image and another's, and on Leonardo, a robot who has been trained to distinguish between his own and other's beliefs, which has important implications for forming moral judgments. Both Breazeal and Turkle also work on, and emphasize the implications of, robots' emotional and social skills, as well as the effects of human projection on social interactions with robots (Breazeal, 2002; Turkle, 2004; Turkle et al., 2006). Based on the above discussion, we would emphasize that despite the need to acknowledge the limitations of neuroscience, it will certainly be of great importance for developing robots, for testing attitudes to robots and for modeling and testing models of cognition and moral capacities; with interdisciplinary relations advantageous to both neuroscientific and engineering disciplines.

Ishiguro's laboratory works not only on the development of his famous Geminoid and other androids, but also develops small interactive humanoids and general sensory mechanisms such as skin sensors, omnidirectional cameras, actuators and sensor networks, and uses neuroscientific tests and eye-tracking for assessing the perceived naturalness of robots' movements. Robots in Ishiguro's lab have access to information that humans may not through a series of strategically placed cameras which map social interactions. Particular robotic abilities, as well as lacks, will be fundamental to designing their ethical systems. Although as a base point non-harming may generally be a prerequisite embedded deeply into the system, the building of robotic soldiers designed to kill humans again reflects the diversity at work, as well as reflecting the challenges for, and current deficiencies of, human moral reasoning. The variety of robots currently being produced and intended for a multiplicity of domains and functions, from the robot soldiers to deep sea and space exploration to those being developed as toys, vacuum cleaners and service robots, suggests that while robots intended for forms of social interactions in the dynamic and ever-changing world of human relations, environments and customs will need to be capable of emotional understanding, social cognition and collaborative behaviour, more basic forms or forms not intended for these purposes need not have these skills.

The evidence that the cognitive economy, rather than consisting only of language and information processes, also involves and evolves through interactions between the brain, the body, and the world, are being incorporated into the field of robotics. There is an increasing emphasis on developmental relationships and ecological control, in which features of the world or semi-autonomous components are employed to offset the need for micro-management and central control systems (MacDorman and Cowley, 2006; Clark, 2007). The distribution and offloading of cognitive processes in a "subsumption architecture" that uses the world as its model, should also include consideration of how these processes could involve ethical systems that are extended (Brooks, 2002).

Floridi and Sanders have proposed a potential step forward for roboethics in terms of rethinking the legal concepts applied. They suggest using the concept of "moral accountability", rather than that of responsibility, in order to sidestep issues of whether robots have person-like capacities of agency. The focus on the responsibility of individual human-like agents has, they suggest, obstructed recognition of the extent to which "distributed morality" is already in operation: "a macroscopic and growing phenomenon of global moral actions and collective responsibilities resulting from…systemic interactions among several agents at a local level" (2004). As neuroethics is considering questions of what levels of observation are appropriate for deciding legal questions of accountability, roboethics and the legal system itself will also face these issues in relation to robotic systems and the increasingly non-trivial causal spread of accountability; *both of accountability in itself and in the way we need to come to understand it*.

Overcoming anxiety: interdisciplinary relations & public communication

> Our responsibility begins with the power to imagine...Turn this on its head and you could say that where there's no power to imagine, no responsibility can arise. (Murakami, 2005)

The authors also wish to draw attention to potential contributions from the history of ideas, since philosophical, historical and literary disciplines all have relevance for the construction of ethical approaches, both in relation to the field of robotics and to our fundamentally *and* increasingly hybrid human forms. Wider spectrum approaches could achieve improved ethical standpoints. The opening of disciplinary doors invites insight into the epistemological assumptions and structures within which we work (our disciplinary "bodies") and can invite less specialized dialogue that is therefore more accessible to public understanding generally. If granted that human subjects' ethical disposition lies in part in their interrelationality, then less closed systems and more interactions between disciplines are potentially advantageous. Besides, peer review systems involve restricted networks of established participants, which do not necessarily invite differing views from those which are the established or favored norms. This is another reason that it is important for researchers to keep channels of dialogue open between disciplines and towards the wider public. Furthermore, issues about the accountability of research relate to broader questions of accountability, in that both disciplines and individuals need to be able to acknowledge their own areas of opacity and inconsistency to the extent that these are cognizable and communicable, rather than eliding them, and to be aware of the already necessarily existing connections both between individual and disciplines.

At "Involving Interface" Tadashi Kobayashi warned of the way in which problems with inflated reporting of research (due to bidding for grants) leads to a crisis of public confidence in science; this fuelling of fears by researchers' exaggerated accounts being then further fuelled by increasingly ungrounded media reporting. The authors suggest that humans' ongoing fear and fascination with technology is also fundamentally due to awareness of our reliance on it. This fear and fascination for present purposes specifically tends to manifest in concerns about the types of robots, technological and biological systems that may be created. While the particular innovations and advances are new they often relate to issues that are familiar from the past, since while we are increasingly discovering revolutionary new ways to extend and transform the biological structure itself, the mind and the subject have always involved extensions into and transformations by the world in which they exist.

Within fictional as well as academic accounts the need to balance the drive to tell a coherent story and the true complexity and richness of detail is an underlying issue. Isaac Asimov, Philip K. Dick, Bernard Wolfe, Richard Powers, Ian McEwan: these are but a few of the writers whose fictional works have in recent decades explored the ethical implications of current and speculated future technological and scientific advances. Hideaki Sena, who combined his own scientific and literary interests in the novel *Parasite Eve*, emphasizes the continuing reciprocity of relations between scientists and novelists, and specifically discussing this in relation to the Japanese context he stresses: "We may be able to gain a realistic view of the environment for robots in Japan by thinking of robot stories as interfaces between culture and science. Images are being passed back and forth between fiction and real-life science, and these two realms are closely interconnected" (2003). Fictional representations of technological hybrids and futuristic scenarios playfully expose and explore the types of fear and fascination aroused by the capacities of robots, cyborgs and human prostheses. Literature's significance is again recognized in Richard Gregory's statement that: "fiction is the look-ahead of Mind that has created the Science in which we find ourselves" (1981).

A final response to fears that once robots eventually develop beyond human capabilities and become able to reprogram themselves, it may not be in a robot's own perceived interests to safeguard human interests, comes not from optimists who trust that human friendly attitudes will survive. Donna Haraway argues that the aim should be to consider both the positive and negative aspects of the possible outcome of our technological developments: "The political struggle is to see from both perspectives at once because each reveals both dominations and possibilities unimaginable from the other vantage point" (1991). Double or even the multiple perspectives provided by diverse art forms and disciplinary bodies are needed.

Yet, there is room for a different type of hopefulness. Our response is motivated by the belief that at such a time of robotic sophistication human cognitive and ethical capabilities would therefore also be likely to outstrip our current forms. Neuroscientific evidence suggests how intimate our interfaces with robots or forms of technological prostheses could be and neural constructivism depicts experience as productive of new neural growth. Clark highlights the significance of this plasticity: "This symbiosis of brain and cognitive technology, repeated again and again but with new technology sculpting new brains in different ways, may be the origin of a golden loop, a virtuous spiral of brain/culture influence that allows human minds to go where no animal minds have gone before" (2001). By these mechanisms human and robotic ethical capabilities could evolve into forms extended beyond the limits of our current imaginations: "Human intelligence is very largely Artificial Intelligence, and even our hopes and fears (and our moral commitments, for they are set by possibilities of achievement) are largely set by existing technology" (Gregory, 1981).

Conclusion

An extended mind approach acknowledges the role of both scientific and sociocultural categories within the ethical context. While certain scientific categories such as genes may be classed as pre-existing, they are like their social counterparts nevertheless subject to the history and evolution not only of the individual, the species, and the world, but also of the terminologies and frameworks used to describe them. Scientific and sociocultural categories are both representative of the world and historical. This hybridity suggests that an integrated equilibrium between top-down and bottom-up approaches will be necessary in creating complexly capable robotic moral agents. This also invites engagement between the arts, sciences and technical disciplines and an acknowledgement of their shared ground as well as their differences: both utilise and interrogate accepted conceptual categories, and dare to imagine new forms, terminologies, and frameworks for thinking about the mind and being human. In concert, they offer fruitful multiple perspectives on these issues.

Our descendents, of all materialities, will hopefully be less biologically bound than we are. Nevertheless, not only our words but also our bodies make evident the extendability and adaptability of humans. The positing of dynamically reciprocal relations between the construction of objects and the construction of subjects logically follows through by positing ethical as well as technological consequences, and as with Butler but on an even wider basis, we contend that the elision of human relationality and materiality, or of our multiplicity and mutability, will not enable but instead undermine the emergence of viable human or robotic ethical models.

Acknowledgments

"Involving Interface" was organized by the authors, and was a satellite session of both the "31st Japan Neuroscience Society Conference" and the "10th Simulation of Adaptive Behaviour Conference". It was also supported by *UK-JAPAN 2008,* organized by the British Embassy and the British Council. The authors especially wish to thank Minoru Asada for chairing this event and the session speakers: Ron Chrisley, Takashi Ikegami, Tadashi Kobayashi, Giacomo Rizzolatti, and Hideaki Sena. Special thanks are also due to Atsushi Iriki of RIKEN BSI and to Hugh Oliphant and Steve Adams at the British Council. Miranda Anderson's participation was supported by a fellowship from the *Japan Society for the Promotion of Science*. Thanks to Sophie Rietti for helpful comments on a draft of this paper. Tamami Fukushi's current affiliation is the Center for Research and Development Strategy, Japan Science and Technology Agency.

References

Agloti, S.M., Cesari, P., Romani, M. and Urgesi, C. (2008). Action Anticipation and Motor Resonance in Elite Basketball Players. *Nature Neuroscience* 11(9): 1109-16.

Anderson, M. (2007a). Early Modern Mirrors. *The Book of the Mirror*. Ed. M. Anderson. Newcastle: Cambridge Scholars Publishing.105-120.

Anderson, M. (2007b). Chaucer and the Subject of the Mirror. *The Book of the Mirror*. Ed. M. Anderson. Newcastle: Cambridge Scholars Publishing. 70-79.

Berti, A. and Frassinetti, F. (2000). When far becomes near: re-mapping of space by tool use. *Journal of Cognitive Neuroscience* 12: 415-20.

Brooks, R. (2002). Does Artificial Life have a Mind? Retrieved on December 15, 2006, from http:www.abc.net.au.

Breazeal, C. (2002). *Designing Sociable Robots*. Cambridge, MA: MIT Press.

Butler, J. (2001). Giving an Account of Oneself. *Diacritics* 31: 22-40.

———. (2005). *Giving an Account of Oneself*. New York: Fordham University Press.

Calvo-Merino, B., Glaser, D.E., Grèzes, J.R., Passingham R.E. and Haggard, P. (2005). Action Observation and Acquired Motor Skills: An fMRI Study with Expert Dancers. *Cerebral Cortex* 15(8): 1243-49.

Clark, A. (1996). Connectionism, moral cognition, and collaborative problem solving. *Minds and Morals. Essays on Ethics and Cognitive Science*. Ed. L. May, M. Friedman and A. Clark. Cambridge, MA: MIT Press.

———. (1997). *Being There: Putting Brain, Body and World Together Again*. Cambridge, MA: MIT Press.

———. (2001). *Mindware*. Oxford: Oxford University Press.

———. (2003). *Natural-Born Cyborgs*. Oxford: Oxford University Press.

———. (2004). We Have Always Been…Cyborgs: Author's Response. *Metascience* 13: 169-81.

———. (2007). Soft Selves and Ecological Control. *Distributed Cognition and the Will: Individual Volition and Social Contex*t. Ed. D. Ross, D. Spurrett, H. Kincaid and G.L. Stephens. Cambridge, MA: MIT Press. 101-22.

———. (2008a). Pressing the Flesh: A Tension in the Study of the Embodied Mind? *Philosophy and Phenomenological Research* 76.1 (2008): 37-59

———. (2008b). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.

Citri, A. and Malenka, R.C. (2008). Synaptic Plasticity: Multiple Forms, Functions, and Mechanisms. *Neuropsychopharmacology* 33: 18–41.

Clark, A. and Chalmers, D. (1998). The Extended Mind. *Analysis* 58: 7-19.

Damasio, A. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Putnam.

Ehrsson, H.H. (2008). The Experimental Induction of Out-of-Body Experiences. *Science* 317: 1048.

Farah, M. and Heberlein, A.S. (2007). Personhood and Neuroscience: Naturalizing or nihilating? *American Journal of Bioethics Neuroscience* 7: 37-48.

Floridi, L. and Sanders, J.W. (2004). On the Morality of Artificial Agents. *Minds and Machines* 14(3): 349-79.

Fukushi, T., Sakura, O. and H. Koizumi. (2007). Ethical Considerations of Neuroscience Research: The Perspectives on Neuroethics in Japan. *Neuroscience Research* 57: 10-16.

Gillet, G. (2007). Cyborgs and Moral Identity. *Journal of Medical Ethics* 32: 79-83.

Gregory, R. (1981). *Mind in Science*. Cambridge: Cambridge University Press.

Gallese, V., Keysers, C. and Rizzolatti G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences* 8(9): 396-403.

Haraway, D.J. (1991). *Simians, Cyborgs, and Women: The Reinvention of Nature*. New York: Routledge.

Hayles, N.K. (2005). *My Mother Was a Computer*. Chicago: University of Chicago Press.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Iacobini, Marco. (2008). Mesial Frontal Cortex and Super Mirror Neurons. *Behavioral and Brain Sciences* 31: 30.

Ishiguro, H. (2005). Android Science: Towards a New Cross-Disciplinary Framework. Paper presented at CogSci-2005 Workshop: Towards Social Mechanisms of Android Science.

Juarrero, A. (2004). We Have Always Been…Cyborgs. *Metascience* 13: 149-53.

Kanda, T. and Ishiguro, H. (2006). An approach for a social robot to understand human relationships. *Interaction Studies* 7(3): 369-403.

Kosslyn, S. (2005). Interview on *Edge*. Retrieved on June 18, 2005, from http://www.edge.org/.

———. (2006). On the Evolution of Human Motivation: The Role of Social Prosthetic Systems. *Evolutionary cognitive neuroscience*. Ed. S. M. Platek and J.P. Keenan. Cambridge, MA: MIT Press. 541-554.

Krach S, Hegel F, Wrede B, Sagerer G, Binkofski F, Kircher T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS One* 3: e2597.

Landecker, H. (2005). Living Differently in Time: Plasticity, Temporality and Cellular Biotechnologies. *Culture Machine*. Retrieved June 6, 2007, from http://culturemachine.tees.ac.uk.

Levy, N. (2008). Introducing Neuroethics. *Neuroethics* 1: 1-8.

MacDorman, K.F. and Cowley, S. (2006). What baboons, babies and Tetris players tell us about interaction: a biosocial view of norm-based social learning. *Connection Science* 18(4): 363-378.

Maravati, A. and Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Science* 8 (2): 79-86.

Mori, M. (1970). Bukimi No Tani [The Uncanny Valley]. *Energy* 7: 33-35.

Murakami, Haruki. (2005). *Kafka on the Shore* [Umibe no Kafka]. Trans. Philip Gabriel. London: Vintage.

Noë, A. (2004). *Action in Perception*. Cambridge, MA: The MIT Press.

Rizzolatti, G. and Craighero, L. (2004). The Mirror Neuron System. *Annual Review of Neuroscience* 27: 169-92.

Rizzolatti, G. and Sinigaglia, C. (2008). *Mirrors in the Brain: How our Minds Share Actions and Emotions*. Trans. Frances Anderson. Oxford: Oxford University Press.

Ross, J. A. (2006). Will Robots See Humans as Dinosaurs? *Journal of Consciousness Studies* 13(12): 97-104.

Sena, H. (2003). Astro Boy Was Born on April 7, 2003. *Japan Echo* 30 (4): 9-12.

Sloman, A., Chrisley, R., and Scheutz, M. (2005). The Architectural Basis of Affective States and Processes. Ed. J.M. Fellous, and Arbib, M.A. *Who Needs Emotions? The Brain Meets the Robot*. Oxford: Oxford University Press. 203-44.

Stelarc. (2005). Prosthetic Head: Intelligence, Awareness, and Agency. *CTheory*. Retrieved on June 19, 2007, from http://www.ctheory.net.

Tsakiris M. (2008). Looking for Myself: Current Multisensory Input Alters Self-Face Recognition, PloS One 3(12): e4040.

Turkle, S. (2004). Whither Psychoanalysis in computer culture? *Psychoanalytical Psychology, American Psychology Association* 21: 16-30.

Turkle, S., Taggart, W., Kidd, C.D. and Dasté O. (2006). Relational artifacts with children and elders: the complexities of cybercompanionship. *Connection Science* 4(4): 347-361.

Turner, J.S. (2000). *The Extended Organism: The Physiology of Animal-Built Structures*. Cambridge, MA: Harvard University Press.

Wallach, W. and Allen, C. (2009). *Moral Machines: Teaching Robots Right from Wrong*. Oxford: Oxford University Press.

Wheeler, Michael. (2005). *Reconstructing the Cognitive World*. Cambridge, MA: MIT Press.