
Book Section:
Prescott, T.J. (2021) Robot self. In: Ang, M.H., Khatib, O. and Siciliano, B., (eds.) Encyclopedia of Robotics. Springer, Berlin, Heidelberg. ISBN: 9783642416101.

https://doi.org/10.1007/978-3-642-41610-1_205-1

Robot Self

Tony J Prescott
Department of Computer Science,
University of Sheffield
Sheffield, United Kingdom
Email: t.j.prescott@sheffield.ac.uk

Synonyms: Robot self model, robot self-awareness, robot self-other distinction

Citation: Prescott, T. J. (in press). Robot Self. In *The Encyclopedia of Robotics*. Ang, M. H., Khatib, O., and Siciliano, B (ed.). Springer: Berlin.

Definition

Robot self describes the capacity of a robot to sense, perceive and conceive of itself as a distinct and integrated entity, localised in space, enduring in time, and capable of initiating and controlling autonomous action. The notion of a robot self is often modelled on psychological theories of the human self that have distinguished different aspects of self—situated, agential, spatiotemporal, interpersonal and conceptual, and have differentiated between the notion of a *core*, or *minimal*, self that is bounded, situated and acts to maintain itself, and an *extended* self, the latter having awareness of itself as an entity with a past and a future, and of social others as also being selves.

Overview

Sense of self is a central phenomenon of human experience and is widely discussed in the psychological and philosophical literatures. It is therefore natural that in developing robots, particularly when following a biomimetic route (see Cangelosi, Tani, this volume), researchers have considered whether it might be possible and useful for robots to have, or acquire, a sense of self. This article considers the nature of the self, whether, and how, a sense of self could be created for robots, and some of the potential societal and ethical implications of developing robots that have sense of self.

What is the self?

There is no agreed consensus as to the nature of self in humans. In philosophy the concept of self is often discussed alongside experiential aspects of the human condition—sentience and consciousness—where the self is seen as being the entity or process that is the *subject* of experience (e.g. Strawson 1997; Sideris et al. 2011). However, psychological analyses of the self have typically taken a broader path (e.g. James 1892/1961; Neisser 1988; Damasio 2010; O'Regan 2011), distinguishing between *self-as-subject*, the experiencer of consciousness, and *self-as-object*, that is, as something we perceive, have knowledge about and act upon. From the perspective of robotics this distinction allows us to think about the self, not purely in terms of the problem of consciousness, and related questions about strong AI, but with regard to the different kinds of self-knowledge that humans acquire and how they give rise to the construction of the self during human development. Providing a robot with access to these forms of self-knowledge could provide the first steps towards creating a robot sense of self.

Treating the self as something that is perceived and constructed first, and as the subject of consciousness second, also makes clear that our understanding of self can draw from many fields of science: developmental and cognitive psychology (Butterworth 1992; Rochat 2019; Neisser

1994/2010), neuroscience, neuropsychology, and evolutionary neurobiology (Feinberg and Keenan 2005; Damasio 2010; Fabbro et al. 2015), linguistics (Lakoff 1996), and anthropology (Cohen 1994). This rich resource of data for understanding the different aspects of self, and their cognitive and neurological bases, can underpin modelling approaches, including robotics, that operationalize and test theories of the self (O'Regan 2011; Prescott and Camilleri 2019; Tani and White 2020; Schürmann et al. 2019; Hafner et al. 2020).

Table 1 illustrates some of the psychological phenomena that we might associate with self-as-object, categorised according to the different aspects of self-knowledge. This scheme, which builds on the taxonomy developed by the cognitive psychologist Ulric Neisser (1988), indicates a possible approach to the deconstruction of the human self and its reconstruction in robotics. Evidence from psychology and neuroscience suggests that the different components of self, identified in this scheme, may be at least partly dissociable as distinct cognitive/neural sub-systems (Prescott 2015; Prescott and Camilleri 2018).

Phenomena of self	**Component of self**
Sensing the body Distinguishing self from the world Having a point-of-view	Situated
Having emotions, drives and motivations Actively seeking information Selecting actions that generate integrated behavior Knowing what events you caused in the world	Agential
Having awareness of where you are Having awareness of a personal past and future Conceiving of the self as localized in space and persisting in time	Spatiotemporal
Learning by imitation Sharing attention Seeing others as selves Having theories of other minds Seeing yourself as others see you	Interpersonal
Having beliefs about who you are (a self-concept) Having personal goals Having a life story (a personal narrative)	Conceptual
Having experience *Having a feeling of being something* *Having a unitary stream of consciousness* *Having a sense of choice (will)* *Having a feeling of being the same thing over time*	*Private*

Table 1. Some psychological phenomena of self and how these might be grouped into different components of self-knowledge based on (Neisser 1988, 1994/2010, 1995) and (Prescott and Camilleri 2019). The private self is shown in italics since these are all first-person phenomena that have an experiential character associated with the self as subject. Although this scheme is based on analysis of the human self it can be considered as a guide for creating a robot sense of self (Prescott and Camilleri 2019).

The question of how we might evidence and measure self-as-subject in humans or implement it in artefacts, remains largely unresolved; ideas based around complexity theory are popular but controversial (Cerullo 2015). A growing number of recent proposals that are based on principles of predictive processing and active inference (Seth and Tsakiris 2018; Woźniak 2018; Newen 2018; Tani and White 2020) attempt to cross, or at least blur, the divide between self-as-subject and self-as-

object. However, an explanatory gap to first-person experience remains. Since at least Hume (1740), and sometimes influenced by ideas from Eastern meditative traditions, there has been a growing stream of thought arguing that there is no self-as-subject, and that our belief in it is a form of illusion (e.g. Metzinger 2009; Sideris et al. 2011; Dennett 1991; O'Regan 2011). One version of this view, that avoids a potentially infinite regress of selves inside selves, is that the embodied mind/brain, considered as a system, is the experiencer, with no single sub-system specifically privileged as the locus of subjectivity.

Key Research Findings

We will first explore the nature and origins of the psychological phenomena associated with the different aspects self-as-object (table 1), and then consider how these have been instantiated in robotic systems.

Multiple aspects of the human self are already in evidence at birth (Stern 1985; Butterworth 1992; Rochat 2019, 2001), while others emerge during the course of infancy and childhood. Those that are manifest in infancy include:

The situated self. The new-born infant is situated in—and has a point-of-view on—the world, has a sense of its own body, and can distinguish sensory signals arising from within its physical boundary from those derived from outside of that boundary—a self-other distinction (Rochat 2001).

The agential self. The situated infant actively orients its view to explore the world and fixate points of interest. Homeostatic and motivational systems, including those related to the maintenance of the self, are linked to the capacity to initiate and express actions that reduce drives and fulfill needs. Knowledge of agency arises from the ability to distinguish events caused by the self from those that have come about independent of self-action (Jeannerod 2003).

These two elements of self can be considered to constitute the *core self***,** also referred to as the *minimal self* (Dennett 1989; Zahavi 2008; Gallagher 2000); for Neisser (1995), they constitute what he termed the *ecological self*. Note that the core self, is not the same as the whole organism, rather, as Dennett has noted it amounts to "the existence of an organization which tends to distinguish, control and preserve portions of the world, an organization that thereby creates and maintains boundaries" (Dennett 1989, p. 167). Importantly, the presence of a core self does not imply or require self-awareness (Bermúdez 1988; Seth and Tsakiris 2018; Gallagher 2000). It is a self that does not know it is a self. The ability to anticipate needs before they arise, implies multiple levels of regulatory control (Sterling 2012; Seth and Tsakiris 2018). The different dynamics, and slower timescales (or stable cycles), of sensed body in comparison to the sensed world, captured by a predictive self-model (see, e.g. Asada et al. 1999; Seth and Tsakiris 2018), could be taken to mark the early emergence of a sense of self.

The sub-cortical brain systems that support the core self are well-developed at birth (LeDoux 2012; Rochat 2019). A core self of some description may also be present in many animal species with a central nervous system. Indeed, based on similarities in neuroanatomy, a core self, similar to our own, is likely to be present in other vertebrates (Fabbro et al. 2015). It is difficult to discern to what extent the self of a human new-born extends beyond this core self, however, in adult humans, the self is a far more complex entity and, almost as a defining characteristic, includes *reflexivity*, that is, the awareness of being (or having) a self.

In terms of the scheme in table 1, the adult human self includes the following components which can also considered to be as aspects of the *extended self*.

The spatiotemporal self. This aspect of self reflects a capacity for self-localisation in both space and time, and the ability to conceive of the self in other times or places. This contrasts with the minimal self which is always situated in the here and now. For example, the capacity for episodic or autobiographical memory allows the self to mentally "time travel" and relive past experiences (Tulving 1985; Prescott et al. 2019). Travelling forward in time is also possible by imagining future scenarios involving the self and how they might evolve. Just as we are able to mentally travel in time, we can also do so in space to imagine ourselves at different locations (Ciaramelli et al. 2010). In mammalian brains these capabilities involve the extended hippocampal system, suggesting that some common brain substrates underlie both the spatial and temporal aspects of self-awareness (Hasselmo 2012). Loss of the capacity for episodic memory leaves the self "marooned in present" (Tulving 1985), while difficulty in localising the self can lead to a feeling of "homelessness" (Norberg 2019). Able to move forward and backward in time, and from place to place, whilst remaining largely unchanged, provides the self with evidence of its own persistence and relative constancy.

The interpersonal self. The newborn human is a highly social animal, adapted to bond rapidly with its caregivers. Aspects of the interpersonal self are therefore present from birth (Stern 1985; Neisser 1995), these gain further traction with the emergence of shared attention, social referencing (looking to adults to understand events), and imitation learning. All of these capacities are present or emerge in the first year of life (Nelson 2007). By age two, most children are able to pass the mirror-test (Amsterdam 1972)—recognizing that the child in the mirror as themself and not another child. This can be taken as evidence of an emerging reflective self model (the ability to see yourself as others see you). During the third year of life, the child comes to understand that other people have different beliefs and knowledge to their own and that they see the world from a different point-of-view; a landmark in human development described by psychologists as the achievement of "theory of mind" (ToM) (Doherty 2009). Theories of the interpersonal self that incorporate ToM often draw on evidence of "mirror neuron" systems that allow the brain to repurpose internal self-models (such as those deployed to represent the child's own physical body) to draw inferences about the actions, beliefs and intentions of others (Schulkin 2000). Growing awareness of other minds might also bring about a sharpening of the insight that one's own perceptions and thoughts are not available to others. In other words, the onset of ToM could provide an ontogenetic landmark for the emergence of a private self (Neisser 1994/2010).

The conceptual self. The final element of self-as-object is a corpus of knowledge about the self—a set of beliefs about who you are, your preferences, goals and so on, and a life story (Neisser 1988). This idea of the self develops gradually through childhood, extrapolated from linguistic exchanges and autobiographical memory, but shaped by cultural norms. Prior to the early school years, children have difficulty providing a narrative account of experience (Bauer 2012), but as we grow up we become more practiced at retelling life events, consolidating the more important ones as part of our life story. For some philosophers (e.g. Dennett 1992; Schechtman 2011; Dennett 1991), this "narrative self"—the central character in the evolving tale that we tell to ourselves and others—is a key element of what we regard as the self. Culture also impacts on the construction of the conceptual self, leading to the widespread adoption of specific beliefs (memes) about the self, such as those concerning its possible immaterial or immortal nature (Leone 2013).

The self system. If the self has all these different aspects and types of knowledge, what brings them together to create a unified sense of self? One answer, that is consistent with the emerging ideas on predictive processing (Nagai, this volume), is that, by constructing/inferring an internal model of a situated, bounded, acting and persistent self, the mind is able to make sense of these otherwise disparate sets of perceptions and memories. This self model can be thought of as the inference to a set of latent variables (hidden factors with predictive utility), and of values of these variables, that reliably distinguish self from other. Just as the mind constructs the world in order to make meaning

from sense data, it also constructs a self in this world, in order to explain its own apparent continuity, capacity for agency, and localised point-of-view. The disintegration of the self in neurological disorders (Feinberg and Keenan 2005) also points to the multi-faceted nature of self, and the existence of processes that create and maintain its integration.

Robot control architectures for instantiating a sense of self

A systems approach suggests that different processes within a robot control system can operate to instantiate the various aspects of self just described, and that the integration of these sub-systems with each other, and with broader perceptual and cognitive capabilities, could give rise to, and project the appearance of, a coherent and unified robot self (Prescott and Camilleri 2019).

Instantiating some aspects of the core self could be a useful place to start (Hafner et al. 2020). Amongst these, an important target would be to demonstrate a robust self-other distinction. Unlike animals, robots typically have limited capacity for interoception, and usually lack full-body tactile sensing. Building a strong sense of the situated self for a robot, that has sensed extent and boundaries, is therefore a non-trivial challenge. Nevertheless, there is a growing field of research on robot learning of body schema and representation of nearby "peripersonal" space (see Hoffmann, this volume). Multisensory integration of sensory flows related to the self (for instance, those arising from vision, touch and proprioception), is another important target (Schürmann et al. 2019). Further interesting work in this direction includes the use of self-touch (for instance, touching a robot finger against the robot's body) as a means of distinguishing self from non-self (Gama et al. 2020). Awareness of body morphology can emerge through "motor babbling" in a robot with suitable joint encoders and accelerometers (Saegusa et al. 2009), a behavior that is also seen in human infants.

Arguably, a robot that has significant autonomy of decision-making, and selects actions according to internal state variables, including those related to self-maintenance and preservation, already possesses an agential self. Many robots have been built that instantiate autonomous agency in this form. A predictive capacity, that can determine the likely sensory consequences of action, can allow a robot to detect its own agency, and lead to the inference of an agential self (Bechtle et al. 2016). Researchers in enactive cognition (e.g. Froese and Ziemke 2009; Di Paolo 2005) have argued for a distinction between robotic agency, and forms of natural agency that arise in organisms that must continuously metabolise to resist the threat of thermodynamic decay (death). This speaks to the possibility that a robot sense of self will be different from a human (or organismic) sense of self in interesting ways such as the urgency attached to self-preservation.

The construction of an extended sense of self, as described here, can only take place in a robot that possesses a cognitive architecture (see Verschure, this volume) with multiple capacities for perception (of both the body and world), emotion, decision-making, memory, attention and reasoning (Prescott and Camilleri 2019; Moulin-Frier et al. 2018) (see also Sandini, this volume). In addition to a motivational/drive system and the ability to sense and model its own morphology, such a robot would require capacities for episodic and spatial memory (e.g. Dominey et al. 2017; Prescott et al. 2019), the ability to construct models of others' actions, beliefs, desires, emotions and intentions (e.g. Trafton et al. 2013; Scassellati 2002; Devin and Alami 2016; Johnson and Demiris 2005; Moulin-Frier et al. 2018; Tani et al. 2004; Asada 2015), and meta-capacities to direct attention to, and inspect and reason about, the perceptions and memories generated by these systems (Novianto and Williams 2009). To create a conceptual self this architecture should also include capacities for abstraction and distillation of important facts and events from episodic memory (see, e.g. Dominey et al. 2017) that can allow the robot to describe and report on itself.

Examples of Application

Many circumstances in which we might consider deploying robots to operate alongside or in co-operation with people (see, e.g. Okada and Contreras, this volume), would benefit from developing aspects of the robot self as described here. For example, providing a robot with improved models of its own physicality, an element of situated self, would make it safer around people. A spatiotemporal self could provide the ability to remember past interactions with people, and support fresh ones, in a more flexible and human-like way than is possible with conventional databases (Prescott et al. 2019). Similarly, an interpersonal self with the capacity to conceive of, and reason about, the beliefs, goals, and motivations of social others, would provide a significant advance for assistive robots that seek to recognize and support human needs.

In human development, it is typically the case that development of these self-systems, and the different forms of knowledge they entail, proceeds alongside, or ahead of, our communicative capacity. In contrast, today's social robots can generate spoken language output, and sometimes non-verbal expressive behaviors, that appear to approach the capabilities of human adults, but without the underlying self, other and situation awareness that people typically possess (see also Prescott and Robillard 2021; Belpaeme, this volume). This can lead to confusion and disappointment in human users. Levelling up to the point where awareness matches language production capability is one of the most important challenges facing social robotics.

Future Directions for Research

From this brief review, we can conclude that there has been some significant, if patchy, progress in developing system components for robots that can detect or construct various kinds of self-knowledge. If the milestone of a minimal (core) robot self has not already been reached, then we are certainly well on the way towards it. The breadth and complexity of the extended self, including its capacity for significant self-awareness, is such that assembling and integrating the prerequisite components remains a substantial and unsolved challenge. Most likely this can be achieved gradually, with simpler versions of the different elements introduced and gradually improved. It also remains open as to what additional mechanisms would be required to give these different types of self-knowledge their experiential "feel". One possible answer is none—that an embodied, and situated robot, with self-as-object and language capabilities, will be able to report, in a suitably informed way, on its own experience, in a manner that consistent with being the locus of that experience (Dennett 1994; Dennett 1991; O'Regan 2011). This possibility makes the robot, as an embodied self-representing and reflexive agent, the self-as-subject rather than any specific sub-system within it.

The development of robots with a capacity for self-awareness creates some significant ethical issues that deserve consideration. These point in two directions. First, there is the concern that robots that have a high degree of autonomy, intrinsic motivational systems, and awareness of themselves and others, may be more difficult to control (in terms of limiting their behavior to that which is useful and good for humans). The development of the artificial self lies on the path towards "artificial general intelligence", which is arguably a riskier endeavor than creating AIs that are focused on narrow domains of expertise (see, e.g. Bostrom 2014). Second, there is the concern that endowing robots with some of the different elements of a sense of self could lead them to have psychological capabilities, such as the ability to suffer (see Asada, Robot Pain and Empathy, this volume), that would make them subjects of ethical concern in their own right (Prescott 2017). Several writers have cautioned against developing robotics/AI in this direction for this reason (Metzinger 2009; Bryson 2018), while others have argued that the state-of-the-art is such that robot rights are already a concern (Gunkel 2018). A responsible approach to research on the robot self should therefore include appraisal of progress in the field with respect to these emerging risks.

Acknowledgements
Development of this contribution was supported by the *Wellcome Trust* "Imagining Technologies for Disability Futures" project [214963/A/18/Z] and by the *EU Horizon 2020 programme* through the *FET Flagship Human Brain Project* (HBP-SGA3, 945539).

References

Amsterdam B (1972) Mirror self-image reactions before age two. Developmental Psychobiology 5 (4):297-305

Asada M (2015) Development of artificial empathy. Neuroscience Research 90:41-50. doi:https://doi.org/10.1016/j.neures.2014.12.002

Asada M, Uchibe E, Hosoda K (1999) Cooperative behavior acquisition for mobile robots in dynamically changing real worlds via vision-based reinforcement learning and development. Artificial Intelligence 110 (2):275-292. doi:https://doi.org/10.1016/S0004-3702(99)00026-0

Bauer PJ (2012) The life I once remembered: the waxing and waning of early memories. In: Berntsen D, Rubin DC (eds) Understanding Autobiographical Memory. CUP, Cambridge, pp 205-225. doi:https://doi.org/10.1017/CBO9781139021937.016

Bechtle S, Schillaci G, Hafner VV On the sense of agency and of object permanence in robots. In: Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob), 19-22 Sept. 2016. pp 166-171. doi:10.1109/DEVLRN.2016.7846812

Bermúdez J (1988) The Paradox of Self-Consciousness. MIT Press, Cambridge, MA

Bostrom N (2014) Superintelligence: Paths, Dangers, Strategies. Oxford University Press, Oxford

Bryson JJ (2018) Patiency is not a virtue: the design of intelligent systems and systems of ethics. Ethics and Information Technology 20 (1):15-26. doi:10.1007/s10676-018-9448-6

Butterworth G (1992) Origins of Self-Perception in Infancy. Psychological Inquiry 3 (2):103-111. doi:10.1207/s15327965pli0302_1

Cerullo MA (2015) The Problem with Phi: A Critique of Integrated Information Theory. PLOS Computational Biology 11 (9):e1004286. doi:10.1371/journal.pcbi.1004286

Ciaramelli E, Rosenbaum R, Solcz S, Levine B, Moscovitch M (2010) Mental space travel: damage to posterior parietal cortex prevents egocentric navigation and reexperiencing of remote spatial memories. Journal of experimental psychology Learning, memory, and cognition 36 3:619-634

Cohen A (1994) Self Consciousness: An Alternative Anthropology of Indentity. Routledge, London

Damasio AR (2010) Self Comes to Mind: Constructing the Conscious Brain. Vintage Books, London

Dennett DC (1989) The Origins of Selves. Cogito 3:163-173

Dennett DC (1991) Conciousness explained. Penguin, London

Dennett DC (1992) The self as the center of narrative gravity. In: Kessel FS, Cole PM, Johnson DL, Hakel MD (eds) Self and Consciousness: Multiple Perspectives. Lawrence Erlbaum, New York,

Dennett DC (1994) The practical requirements for making a conscious robot. Philosophical Transactions of the Royal Society of London A 349:133-146

Devin S, Alami R An implemented theory of mind to improve human-robot shared plans execution. In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 7-10 March 2016 2016. pp 319-326. doi:10.1109/HRI.2016.7451768

Di Paolo EA (2005) Autopoiesis, Adaptivity, Teleology, Agency. Phenomenology and the Cognitive Sciences 4 (4):429-452. doi:10.1007/s11097-005-9002-y

Doherty M (2009) Theory of Mind: How Children Understand Others' Thoughts and Feelings. Psychology Press, Hove

Dominey PF, Paléologue V, Pandey AK, Ventre-Dominey J Improving quality of life with a narrative companion. In: 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 28 Aug.-1 Sept. 2017 2017. pp 127-134. doi:10.1109/ROMAN.2017.8172291

Fabbro F, Aglioti SM, Bergamasco M, Clarici A, Panksepp J (2015) Evolutionary aspects of self- and world consciousness in vertebrates. Frontiers in human neuroscience 9:157-157. doi:10.3389/fnhum.2015.00157

Feinberg TE, Keenan JP (2005) The Lost Self: Pathologies of the Brain and Identity. OUP, Oxford

Froese T, Ziemke T (2009) Enactive artificial intelligence: Investigating the systemic organization of life and mind. Artificial Intelligence 173 (3):466-500. doi:https://doi.org/10.1016/j.artint.2008.12.001

Gallagher S (2000) Philosophical conceptions of the self: Implications for cognitive science. Trends in Cognitive Sciences 4 (1):14–21

Gama F, Shcherban M, Rolf M, Hoffmann M Active exploration for body model learning through self-touch on a humanoid robot with artificial skin. In: 2020 Joint IEEE 10th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob), 26-30 Oct. 2020 2020. pp 1-8. doi:10.1109/ICDL-EpiRob48136.2020.9278035

Gunkel D (2018) Robot Rights. MIT Press, Cambridge, MA

Hafner VV, Loviken P, Pico Villalpando A, Schillaci G (2020) Prerequisites for an Artificial Self. Frontiers in Neurorobotics 14:5

Hasselmo ME (2012) How we Remember: Brain Mechanisms of Episodic Memory. MIT Press, Cambridge, MA

Hume D (1740) A Treatise on Human Nature.

James W (1892/1961) Psychology The Briefer Course. Harper, New York

Jeannerod M (2003) The mechanism of self-recognition in humans. Behavioural Brain Research 142 (1):1-15. doi:doi.org/10.1016/S0166-4328(02)00384-4

Johnson M, Demiris Y (2005) Perceptual perspective taking and action recognition. International Journal of Advanced Robotic Systems 2 (4):32. doi:10.5772/5775

Lakoff G (1996) Sorry I'm not myself today: The metaphor system for conceptualizing the self. In: Fauconnier G, Sweetser E (eds) Spaces, Worlds, and Grammar. University of Chicago Press, Chicago,

LeDoux J (2012) Rethinking the emotional brain. Neuron 73 (4):653-676. doi:10.1016/j.neuron.2012.02.004

Leone M (2013) Signs of the Soul: Toward a Semiotics of Religious Subjectivity. Signs and Society 1 (1):115-159. doi:10.1086/670169

Metzinger T (2009) The Ego Tunnel: The Science of the Mind and the Myth of the Self. Basic Books, New York

Moulin-Frier C, Fischer T, Petit M, Pointeau G, Puigbo J, Pattacini U, Low SC, Camilleri D, Nguyen P, Hoffmann M, Chang HJ, Zambelli M, Mealier A, Damianou A, Metta G, Prescott TJ, Demiris Y, Dominey PF, Verschure PFMJ (2018) DAC-h3: A proactive robot cognitive architecture to acquire and xxpress knowledge about the world and the self. IEEE Transactions on Cognitive and Developmental Systems 10 (4):1005-1022. doi:10.1109/TCDS.2017.2754143

Neisser U (1988) Five kinds of self-knowledge. Philosophical Psychology 1:35-59. doi:10.1080/09515088808572924

Neisser U (1994/2010) The Perceived Self. Cambridge University Press, Cambridge, UK

Neisser U (1995) Criteria for an ecological self. In: Rochat P (ed) The Self in Infancy: Theory and Research. Elsevier, Amsterdam,

Nelson K (2007) Young Minds in Social Worlds: Experience, Meaning and Memory. Harvard University Press, Cambridge, MA

Newen A (2018) The Embodied Self, the Pattern Theory of Self, and the Predictive Mind. Frontiers in psychology 9:2270-2270. doi:10.3389/fpsyg.2018.02270

Norberg A (2019) Sense of self among persons with advanced dementia. In: Wisniewski T (ed) Alzherimer's Disease. Codon Publications, Brisbane,

Novianto R, Williams M The role of attention in robot self-awareness. In: RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication, 27 Sept.-2 Oct. 2009 2009. pp 1047-1053. doi:10.1109/ROMAN.2009.5326155

O'Regan K (2011) Why Red Doesn't Sound Like a Bell: Understanding the Feel of Consciousness. Oxford University Press, Oxford

Prescott TJ (2015) Me in the Machine. New Scientist.

Prescott TJ (2017) Robots are not just tools. Connection Science 29 (2):142-149. doi:10.1080/09540091.2017.1279125

Prescott TJ, Camilleri D (2018) The Synthetic Psychology of the Self. In: Sequeira J, VenturaI R, Ferraira I (eds) Cognitive Architectures. Springer Verlag., Berlin,

Prescott TJ, Camilleri D (2019) The Synthetic Psychology of the Self. In: Aldinhas Ferreira MI, Silva Sequeira J, Ventura R (eds) Cognitive Architectures. Springer International Publishing, Cham, pp 85-104. doi:10.1007/978-3-319-97550-4_7

Prescott TJ, Camilleri D, Martinez-Hernandez U, Damianou A, Lawrence ND (2019) Memory and mental time travel in humans and social robots. Philosophical Transactions of the Royal Society B: Biological Sciences 374 (1771):20180025. doi:doi:10.1098/rstb.2018.0025

Prescott TJ, Robillard JM (2021) Are friends electric? The benefits and risks of human-robot relationships. iScience 24 (1). doi:10.1016/j.isci.2020.101993

Rochat P (2001) The Infant's World. Harvard University Press, Cambridge, MA

Rochat P (2019) Self-Unity as Ground Zero of Learning and Development. Frontiers in Psychology 10:414

Saegusa R, Metta G, Sandini G, Sakka S Active motor babbling for sensorimotor learning. In: 2008 IEEE International Conference on Robotics and Biomimetics, 22-25 Feb. 2009 2009. pp 794-799. doi:10.1109/ROBIO.2009.4913101

Scassellati B (2002) Theory of mind for a humanoid robot. Autonomous Robots 12 (1):13-24. doi:10.1023/A:1013298507114

Schechtman M (2011) The narrative self. In: Gallagher S (ed) The Oxford Handbook of the Self. Oxford University Press,

Schulkin J (2000) Theory of mind and mirroring neurons. Trends in Cognitive Sciences 4 (7):252-254. doi:10.1016/S1364-6613(00)01500-X
Schürmann T, Mohler BJ, Peters J, Beckerle P (2019) How Cognitive Models of Human Body Experience Might Push Robotics. Frontiers in Neurorobotics 13 (14). doi:10.3389/fnbot.2019.00014
Seth AK, Tsakiris M (2018) Being a Beast Machine: The Somatic Basis of Selfhood. Trends in Cognitive Sciences 22 (11):969-981. doi:10.1016/j.tics.2018.08.008
Sideris M, Thompson E, Zahavi D (2011) Self, No Self? Perspectives from Analytical, Phenomenological and Indian Traditions. OUP, Oxford
Sterling P (2012) Allostasis: A model of predictive regulation. Physiology & Behavior 106 (1):5-15. doi:https://doi.org/10.1016/j.physbeh.2011.06.004
Stern D (1985) The interpersonal world of the infant. Basic Books, New York
Strawson G (1997) The self. Journal of Consciousness Studies 4 (5/6):405-428
Tani J, Ito M, Sugita Y (2004) Self-organization of distributedly represented multiple behavior schemata in a mirror system: reviews of robot experiments using RNNPB. Neural Networks 17 (8):1273-1289. doi:https://doi.org/10.1016/j.neunet.2004.05.007
Tani J, White J (2020) Cognitive neurorobotics and self in the shared world, a focused review of ongoing research. Adaptive Behavior:1059712320962158. doi:10.1177/1059712320962158
Trafton JG, Hiatt LM, Harrison AM, Tamborello FP, Khemlani SS, Schultz AC (2013) ACT-R/E: an embodied cognitive architecture for human-robot interaction. J Hum-Robot Interact 2 (1):30–55. doi:10.5898/JHRI.2.1.Trafton
Tulving E (1985) Memory and consciousness. Canadian Journal of Psychology 26 (1):1-12
Woźniak M (2018) "I" and "Me": The Self in the Context of Consciousness. Frontiers in psychology 9:1656-1656. doi:10.3389/fpsyg.2018.01656
Zahavi D (2008) Subjectivity and Selfhood: Investigating the First-Person Perspective. MIT Press, Cambridge, MA

Cross-references

Belpaeme, T., Interaction
Sandini, G., Cognitive robotics
Cangelosi, A., Developmental robotics
Hoffmann, M., Biologically-inspired robot body models and self-calibration
Nagai, Y., Predictive coding for cognitive development
Okada, H. and Contreras, L., Domestic and personal robots
Tani, J., Neurorobotics
Verschure, P.F.M.J., Cognitive architecture
Asada, M., Robot pain and empathy