This is a repository copy of *Understanding the sense of self through robotics*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/219433/

Version: Accepted Version

Article:
Prescott, T.J. orcid.org/0000-0003-4927-5390, Vogeley, K. and Wykowska, A. orcid.org/0000-0003-3323-7357 (2024) Understanding the sense of self through robotics. Science Robotics, 9 (95). eadn2733. ISSN 2470-9476

https://doi.org/10.1126/scirobotics.adn2733

© 2024 The Authors. Except as otherwise noted, this author-accepted version of a journal article published in Science Robotics is made available via the University of Sheffield Research Publications and Copyright Policy under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Understanding the Sense of Self through Robotics

Tony J. Prescott[1], Kai Vogeley,[2,3] and Agnieska Wykowska,[4]*

[1] Department of Computer Science and Sheffield Robotics, University of Sheffield, Sheffield, UK

[2] Department of Psychiatry, University Hospital Cologne, Cologne, Germany

[3] Institute for Neuroscience and Medicine, Research Center Juelich, Juelich, Germany

[4] Italian Institute of Technology, Social Cognition in Human-Robot Interaction Unit; Genova, Italy

*Corresponding author. Email: agnieszka.wykowska@iit.it

Abstract: Robotics can play a useful role in the scientific understanding of the sense of self, both through the construction of embodied models of the self and through the use of robots as experimental probes to explore the human self. In both cases, the embodiment of the robot allows us devise and test hypotheses about the nature of the self with regard to its development, its manifestation in behavior, and the diversity of selves in humans, animals, and potentially machines. This paper reviews robotics research that addresses the topic of the self—the minimal self, the extended self, and disorders of the self—and highlights future directions and open challenges in understanding the self through constructing its components in artificial systems. An emerging view is that key phenomena of the self can be generated in robots with suitably configured sensor and actuator systems and a layered cognitive architecture involving networks of predictive models.

Cite as: Prescott, Tony J., Kai Vogeley, and Agnieska Wykowska, Understanding the sense of self through robotics. *Science Robotics*, 9(95), p. eadn2733.

Abbreviations

SoA—sense of agency

SoO—sense of (body) ownership

ToM—theory of mind

1. INTRODUCTION

Our lived experience of "me-here-now" is based on phenomena that are constitutive of the self, including feelings of body ownership and of agency, both reliant on a body-centred spatial perspectivity, and a sense of transtemporal unity of experience [1]. These phenomena (see below) anchor our mental life including our thoughts, memories, feelings, and intentions to act. Being at the core of human experience, the nature of the self has been a foundational topic in philosophy and has been brought into the realm of empirical research by psychology and cognitive neuroscience. The experience of self is also profoundly impacted in mental illness, leading to strong interest in the topic in the fields of psychiatry and psychopathology [2]. As we will examine in this article, the self is increasingly explored in cognitive robotics, with growing potential to inform these other realms of scientific endeavour.

We consider three different ways in which robotics can be made useful for the scientific study of the self. First, robotics allows both the systematic construction of the self, by integrating different components in embodied computational models, and its deconstruction, by isolating components at the functional level. Second, robots can be used as sophisticated probes (apparatuses) and interaction partners in experimental protocols that study the human self. Third, and largely unexplored, studies in which robots include a modified, or perturbed, model of self could assist in understanding the diversity of human selves including disturbances of the self. Our main focus will be on research using embodied systems (physical robots), often with a humanoid or zoomorphic form, and not disembodied "bots" such as conversational agents, we will also touch on the contribution of other robotic systems such as prostheses and exoskeletons.

2. THEORIES OF SELF

While Descartes famously considered that the self was simple and distinct from the body, he elsewhere described the self as a composite equated with the whole person [3]. Hume understood the self as a bundle of experiences [4]. This tension between the intuition of the self as an unanalysable essence, and the self as an aggregate enmeshed with the body, is reflected in hundreds of years of debate in science and philosophy about the human condition. For the psychologist, William James, there were two sides to the self—one the subject of experience ("I") the other its object ("me") [5]. While philosophy and consciousness science have largely focused on the problem of subjectivity; psychology, and latterly, cognitive neuroscience have developed a complementary understanding of self as object. For instance, a taxonomy of self was developed by Ulrich Neisser who decomposed different forms of self-knowledge into ecological (situated), interpersonal, temporally-extended, conceptual and private aspects [6]. A further flourishing line of research has focused on the narrative self—the construction of a coherent and meaningful identity through storytelling [7, 8].

2.1. A systems perspective on the self

A contemporary view of the self builds on this earlier research, and on the broader "systems" approach in biology that views organisms, including humans, as complex dynamical systems [9]. Systems are aggregates of many interacting parts, where the system, as a whole, can have emergent properties that are not present in any of its components. From a systems perspective, the self could be constituted by patterns of attractor dynamics in neural activity within widely distributed parts of the brain, closely coupled with non-neural systems throughout the body, and, via verbal and non-verbal communication, with other embodied selves. Such a view aligns with enactivist and embodied cognition views of self-hood [10-12], recognizing the dependency of self on the body, on the wider environmental context, and on the construction of self through interaction with others. Gallagher [13] has proposed a related view of self as a ‘pattern” arising from the activity of multiple self-related processes. Results from developmental psychology, cognitive neuroscience and psychopathology also support the notion of dissociable self sub-systems that collectively give rise to the human sense of self [14].

In this review, we also approach the self as a construct that binds across sub-systems to form an emergent and unified whole constituted by (but not limited to) core phenomena of ownership, agency, and transtemporal unity. These key phenomena are integrated in an embodied self-model instantiated within a cluster of predictive systems established within the brain and body and its ecological niche [15-21].

Ownership is, for example, reflected in the experiential quality of “mine-ness” of my perceptions, feelings, thoughts, and intentions [22], as also expressed through pronominal syntax in language, and as determined by a flexible boundary—the “self-manifold” (see below) [23]. Agency refers to the experience that I am the author or originator of my movements [24, 25] as determined by my capacity to predict their effects or infer my own causal influence on an outcome [26]. Both ownership and agency are organised according to a spatial, body-centred perspective [16, 19, 27].

Beyond the persistence of the body, the basis for transtemporal unity of experience is the existence of a long-term coherent whole of beliefs and attitudes—the narrative self—that extends through lifetime and that stands for “the idea of a single person, a single subject of experience and action” [28] (p. 396). Moreover, this allows one to act as an “observer, agent, and guardian of the continuity of experience” ([29], p. 161). As the self develops, including through interaction with other selves, we do so, increasingly, based on interpretable meanings that develop during those interactions, and more broadly within our culture, and that can change over time [30, 31]. Active engagement with the world and other persons therefore profoundly impacts on the construction and experience of the self.

As we will explore in section 3, we consider that robotics, allied with appropriate AI methods such as generative modelling, can provide a useful testbed to explore this broad hypothesis of the self as both a cluster of dissociable sub-systems and as a constructed whole. By assembling appropriate modules within the cognitive architecture of a robot, a sense of self can emerge as having a unique perspective through its sensors and agency through its actuators. Moreover, such a self, may also infer its own persistence from the predictability and consistency of that embodiment over time.

Whilst we primarily consider psychological aspects of self in this article, this is not to dismiss the importance of understanding how the self becomes realized in the physical and neurobiological substrates of the body, and through the dynamic coupling of the brain and body with the environment [22, 32]. Our position is similar to Harnad’s [33] “robot functionalism” which asserts the importance, and potential primacy, of nonsymbolic capacities in grounding cognitive capacities and sense of self (see also [34]). More broadly, we consider that a full theory of self will encompass multiple levels of explanation [35] and that robotics research on sense of self can be usefully informed by neurobiology, for instance with regard to its realization through layered control architectures (e.g. [36-39]). We note that some research on self asserts that a sense of self, and particularly a capacity for subjective experience, cannot be realized in non-biological substrates. We will return to this important question in section 6.

2.2. Minimal and extended selves

From an evolutionary perspective, biological organisms distinguish processes that occur within the body boundary from those that are external to it. This provides the basis for a distinction between self and other which is one of the most foundational aspect of self [40]. Many organisms also have the capacity to distinguish the consequences of their own actions from the broader flow of environmental events via reafference—the effects of action on what is sensed. For Jékely et al. [41] this provides all bilaterian animals with an elementary sense of self.

Broadly, two aspects of self—sense of body ownership (SoO) and sense of agency (SoA), building on this primary distinction between self and other, constitute the central elements for what, following Dennett [40] and Gallagher [42], we will call the *minimal self*. This also largely coincides with Neisser's notion of the ecological self as "the embodied individual purposefully engaged with the environment" [6] (p. 20). Related proposals have been made by Damasio [43] who describes a "core self"—in which SoA and SoO build on a cluster of "protoself" brain and bodily processes that ensure survival, and by Panksepp [44] who discusses the "primal self"—emphasising the convergence of emotional (evaluative), sensory and motor schema with an integrative body map. Whilst definitions of what constitutes the minimal self may be somewhat fuzzy, there is broad agreement that this elementary form of self needs no capacity to reflect on itself, form attitudes or beliefs about itself, or to conceive or be aware of itself as persisting in time [42].

Insights from developmental psychology suggest the presence of at least a minimal self, and probably more, in the human infant [45, 46]. Over the course of development, the child will construct further aspects of self, including the capacity for mental time travel into the past or future that provides for a sense of self that is extended in time [47], and awareness of itself as located in space [48]. With the achievement of "theory of mind" (ToM) [49]—the ability to infer the mental states of others from their observed behaviour—the child gains an understanding of social others as also being selves, building on what is likely to be a simpler form of awareness of others as social entities and agents that may be present at birth or emerge during the first years (indicated by joint attention, for example). As the child grows older, the ability to recall past episodes leads to consolidation of an autobiographical memory, the construction of a sense of self as the "narrative centre" of its emerging life story [7, 8]. We will refer to these additional aspects of self collectively as the "extended self" [14, 50].

3. ROBOTS AS MODELS OF THE SELF

A variety of robots have been used to investigate different aspects of self and related behavioural phenomena, as illustrated in figure 1. The approach of using robots to model the self follows an "understanding through building" strategy that capitalises on the value of physical models in investigating complex systems [35]. Robots provide testbeds in which candidate self sub-systems, including models of target brain processes, can be embedded as part of a wider control architecture [51] and evaluated in real-world settings that include embodied others. This embedding challenges theoretical proposals to be more fully specified and provides tests of their sufficiency and completeness, particularly with regard to their capacity to generate behavioural phenomena linked to sense of self. Through this process, robots can make more amenable foundational questions about self, including about the nature of the sense of the self, the preconditions for the emergence of sense of self, and the potential diversity of selves, including the minimal case [14].

Theories of self [15, 17-20, 22] are increasingly predicated on a predictive processing view that sees the construction of the self as a process of minimising prediction error in a network of generative models. Specific models may relate to different modalities of experience such as proprioception, interoception, or exteroception, or may construct multimodal latent space representations (across lower-level models) that encode self-related information such the structure, location and pose of the body. At higher levels, such models will encode more abstract concepts such as memories, intentions, goals, beliefs about the self, and ultimately a concept of self. Such as view is particularly amenable to implementation and testing via robotic embodiment [21, 23, 39, 52].

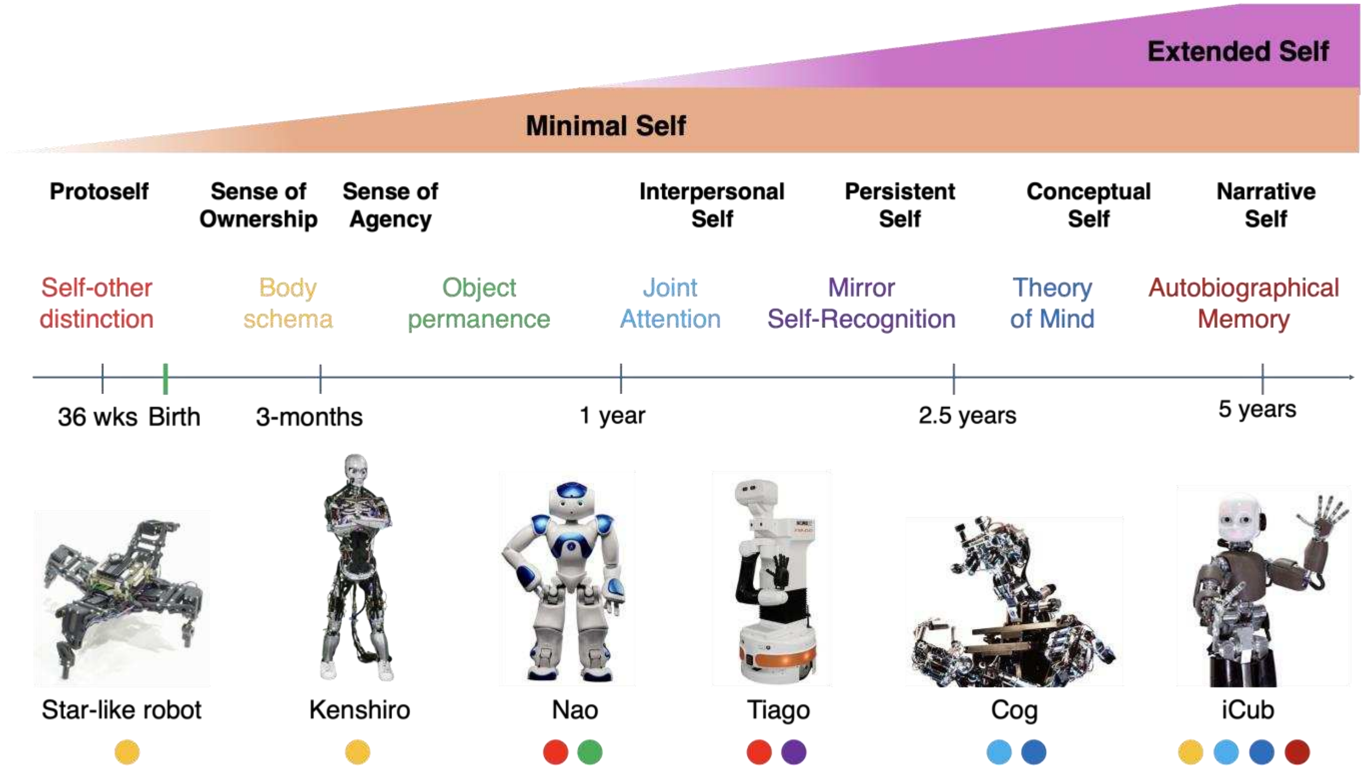

Figure 1. Top. Aspects of self and some of the psychological capacities (coloured text) associated with them, shown above an approximate developmental timeline (log scale) based on [45-47, 53]. In this article we discuss early emerging aspects of self as related to the "minimal self" and later-emerging aspects as relating to the "extended self". **Bottom.** Examples of robots used to investigate different aspects of self through embodied modelling. Coloured circles indicate specific psychological capacities these robots have been used to emulate. From left to right: A star-like robot used to investigate the emergence of a body schema through continuous self-modelling [54]. The *Kenshiro* [55] robot mimics human musculoskeletal structure and is being developed as a platform for simulating the human body schema [56]. Commercial humanoid platforms such as *Nao, Pepper* (not shown) and *Tiago* have been used to investigate phenomena associate with self-other distinction [57, 58] and the emergence of the extended self—object permanence [59], theory of mind [60] and mirror-self recognition [58]. *Cog*, an upper-torso humanoid, designed as a tool to study human cognitive development [61], was used to investigate joint attention [62] and theory of mind [63]. The *iCub* humanoid [64] is a full-body robot designed to emulate the motor and sensory capacities of a human child, it has been used extensively to investigate phenomena of self, including body schema [36, 65-67], joint attention [36], theory of mind [68], episodic/autobiographical memory and the narrative self [36, 39, 69, 70]. See text for further explanation. *Photo credits: star-like robot, Josh Bongard, Victor Zykov, and Hod Lipson; Kenshiro, adapted from [55] with permission from AAAS; Nao, Softbank robotics; Tiago, Pal Robotics, S. L.; Cog, Sam Ogden and Science Photo Library (license obtained); iCub, Italian Institute of Technology.*

3.1. Constructing the minimal self

The minimal self, as described above, is composed of two principal subsystems relating to SoO and SoA. The embodiment of organismic selves further implies a variety of primary capabilities including sensing the body and the environment, having a self-other distinction, and having a perspective on the world (a point-of-view). In human adults these capacities are reflective (i.e. accessible to introspection) and involve meta-cognition, however, behavioural markers indicate that human infants have SoO and SoA but without the adult capacity to conceptualise their experience of self [71]. In developing robot models of the minimal self a key interest is therefore in emulating these pre-reflective forms of ownership and agency.

The importance of sensing the body is broadly recognised in discussions of self but encompasses at least two distinct themes—one focusing on proprioception and kinesthetics [24, 72], and one on interoception, that is, sensing of internal bodily processes [20, 73]. In robotics, kinesthetics and proprioception are widely studied and emulated, particularly with regard to safe control of stance and movement, but also to construct a robot body schema that could act as a model of the human body schema (see [56, 65, 74-76]). Despite a wealth of studies and data there is still a lack of clear understanding of how the brain constructs and uses representations of the body, partly due to the difficulty of recording in the brain in moving animals. Robotics can contribute by embedding models of hypothesised mechanisms in physical systems engaged with real environments. A range of theories about both explicit and implicit encodings of the body schema have been explored using robotic systems some of which utilise robots whose actuator and mechanical systems are closely modelled on human anatomy and physiology [56]. Hypotheses under examination include that latent representations that map across modalities, such as proprioception and touch, could be useful in understanding body-related brain activity in areas such as parietal cortex [76, 77]. More broadly, the capacity to learn lower-dimensional representations of the robot's sensorimotor and action spaces (see for an example [78]), that discover and encode relevant invariances, can create what Hafner has termed the "self-manifold" [23]. By learning this manifold, which is both flexible and reconfigurable, the robot can acquire an understanding of how its body can move and sense in space. This idea also operationalises Damasio's [79] notion of a "sensory portal" as being the configuration of brain and body systems that enable any given sensory capacity.

Whilst studies of exteroceptive and proprioceptive mechanisms in robotics are commonplace, research on interoception is largely lacking—the internal milieu of a robot is much simpler than that of a biological organism as functions such as respiration, digestion, and excretion, that are fundamental to animal life, are not needed. The lack of interoception, which grounds homeostasis and emotion in relation to these bodily processes, and is often discussed in relation to self-hood (e.g. [44]), has been highlighted as a potential weakness of embodied artificial systems [80]. Research on self-healing sensorized materials for soft robots [81, 82], energy harvesting [82], and "embodied energy" [83] (using power generation that are more similar to these used by animals), should lead to more "internally-aware" robotic systems.

The development of a self-other distinction has been widely studied in robotics, examples include the emergence of this distinction through self-touch [75], via visual feedback and the construction of a body model [58, 84, 85], and through the construction and monitoring of peripersonal space [67]. A key boundary of the human self is provided by the skin which provides both exteroceptive and interoceptive signals (such as warmth) while also acting as a channel for both perception (e.g. shape, texture) and valence (e.g. pain, social touch). Whilst there has been significant progress towards developing robotic models of skin [86], existing technologies are generally unimodal and are focused almost exclusively on perception. Few robots have been enveloped in artificial skin, so, in existing systems, this acts as a partial boundary at best. The limitations of sensing at the body-world boundary, together with the relative lack of interoception, means that current robots have a more impoverished representation of their own embodiment than animals, though this may change with the development of soft robots and more energy-autonomous devices [81, 83, 87].

With regard to perspective taking, it is uncontroversial that a physical robot equipped with sensory systems, such as directional cameras and microphones, has a point-of-view in the literal sense of that term. In other words, there is a position in the world that is occupied by the robot and that cannot be occupied directly by anyone or anything else at that time. There is also a specific perspective or view of the world from that position, that is unique to the robot at any moment, and dependent on its sensor configuration. Blanke and Metzinger describe a "weak first person perspective" (weak 1pp) as requiring "a spatial frame of reference, plus a global body representation, with a perspective originating within this body representation" ([88], p.7), a weak 1pp is required for subjective experience, according to these authors, but is not, in itself, sufficient for it. Thus, a suitable-designed robot with something akin to the self-manifold, as discussed above, would qualify as having weak 1pp. Blanke and Metzinger also consider a strong 1pp as emerging when the agent achieves the capacity to reflect on itself and to have a narrative-based identity. This would be a property of the extended self (as discussed further below).

Sense of agency (SoA) is the experience of control over one's own actions and their consequences for external events [72]. SoA is also related to the notion of autonomy—an entity is autonomous to the extent to which its actions are intrinsically determined rather than externally directed. The capacity to interact with the physical world through embodied sensors and actuators provides the fundamental material substrates for embodied agency. Combined with value systems, which could be hardwired or acquired through reinforcement learning, and a self-other distinction (created by body representations as discussed above), robots are beginning to meet the core requirements, suggested by Damasio [79] and Panksepp [44], for grounding SoA in the integration of external and internal sensing with value-based action selection. The capacity to direct attention and to actively sense the world (i.e. to control the sensing apparatus in an information-seeking way) are further agency-related phenomena, that have been related to sense of self (e.g. [88]) and extensively explored robotics [89-91]. Biological agency is also grounded in homeostatic and allostatic regulatory systems [38, 92], a layered control architecture of motivational systems [38] and neuromodulatory systems that are integrated with bodily processes [93]. Models of some of these mechanisms are now being investigated in brain-inspired robotics [36, 94-97].

A key pre-requisite for sense of agency in humans is considered to be the capacity to predict the sensory consequences of intrinsically generated actions and thereby distinguish those consequences from other events [98]. Such capabilities have been developed for robots [59, 99] and applied to emulate milestones in infant development such as object permanence—understanding that objects continue to exist when they are out of sight [59]. While philosophy has suggested the necessity of object permanence for the development of the child's ability to conceive of itself [100], robotics research has shown that a sense of agency may be required for the discovery of object permanence [59]. This illustrates how research on robotics could aid our understanding of the dependencies between different aspects and phenomena of self, and consequently their developmental timeline (as illustrated in figure 1).

The phenomena associated with the minimal self, including SoO, SoA and self-other distinction, are conceptual distinct but closely related. They are not all or nothing, each is likely to depend on multiple substrates some of which may be overlapping. According to Rochat [45, 46] the infant has both a sense of its own body and of its own agency by age 2-3 months, many aspect of these being absent in the newborn, however, "the development of self-knowledge does not start from an initial state of confusion. Infants are born with the perceptual means to discriminate themselves from other objects and appear to use these means to sense themselves as differentiated, situated, and effective in their environment" ([45], p. 108). Robotics offers as means of operationalising, analysing and distinguishing these concepts, their interdependence and time course. For instance, the emergence of a self-other distinction prior to birth has been proposed as deriving, in part, from the sense of touch which develops early during gestation, and the different experience of touching the self, compared to touching the lining of the womb [101]. Simulations of a brain/body model of the human fetus [102] and self-touch studies in robotics have demonstrated the possibility of learning about the layout of the body using somatosensation coupled with proprioception [66, 77]. This can help us to identify the kind of body representations that may already be in place before birth.

3.2. Constructing the extended self

The human self, and likely that of other complex animals, goes significantly beyond what we have described as the minimal self. Multiple phenomena of self that can be grouped into at least three distinct subsystems—transtemporal, interpersonal and narrative—constitute what we will describe as the extended self.

Capacities for localising in time, which relies on episodic memory, have been investigated in the context of cognitive architectures for humanoid robots (e.g. [70, 103, 104], for review see [39]), and provide additional evidence, to support that from neuroscience, that the same mechanisms can support both remembering of past events and imagining of future events. Localisation in time is widely emphasised in theories of self, however, being localised in space is also important for sense of self and involves similar neural substrates including the hippocampal system and default mode network [105]. Capacities for spatial localization are well developed in robotics in the form of simultaneous localization and mapping algorithms (SLAM) and can be used to inform theories of subsystems in the human brain underlying orientation and navigation [106].

Aspects of the interpersonal self that have been explored in robotics include joint attention [36, 62, 107], imitation [108, 109] and theory of mind (ToM), viewed as the capacity to attribute goals, beliefs, desires, emotions and intentions to others and to recognise that these are different from one's own (e.g. [63, 110, 111]). Some robots have been endowed with capacities for abstraction and distillation of important facts and events from episodic memory (see, e.g. [103, 104]) that can allow the robot to describe and report on itself [36, 112] and could form the beginning of a narrative self and acquired self-concept. Finally, there have been robotic investigations of the capacity to direct attention to, and inspect and reason about, the perceptions and memories generated by these systems [91, 113] that could provide the beginnings of self-reflection.

The development of a human-like extended sense of self for robotics will require a cognitive architecture with capacities for perception (of both the body and world), emotion, decision-making, memory, attention, and reasoning. Examples of existing architectures that go in this direction include the "distributed adaptive control" architecture which has been embodied in the iCub humanoid [14, 36] and the cognitive architecture for the Karlsruhe humanoid [104, 114]. Evidence from studies of the human sense of self suggest the need for a layered architecture [38] in which core self sub-systems, implemented at the level of the brainstem and available within the first few months life, are modulated by a hierarchy of predictive sub-systems specified in the forebrain and particularly the cerebral cortices [19, 39].

4. ROBOTS AS EXPERIMENTAL PROBES TO STUDY THE HUMAN SELF

This section focuses on the use of robots as sophisticated "tools" to study the human self. As such, the robots discussed here are not necessarily endowed with the model of the self. They might not even be autonomous or endowed with any sophisticated cognitive architecture. Instead, they can be considered as technology enabling the study of human cognition that augment existing forms of experimental apparatus [115]. Robots offer greater ecological validity than approaches such as presenting stimuli via computers. At the same time, they provide improved experimental control compared to completely naturalistic protocols where participants are tested in interaction with the environment or with others "in the wild" [115-117]. The use of robots as experimental tools depends, of course, on the phenomenon under investigation. Robots can serve as effective proxies for studying social cognition mechanisms because they are embodied and thus can engage participants in tasks such as joint action and joint manipulation of physical objects. Additionally, they can be designed to resemble humans in their appearance and motor repertoire and to provide the impression of a physically-present social agent.

In this section, we will focus on research relating to the cognitive mechanisms underlying sense of body ownership (SoO) and sense of agency (SoA) as core aspects of the minimal self [26] and on mechanisms underlying the inter-personal self as an example of research related to the extended self.

4.1. Understanding the minimal self

Broadly speaking, research on SoO in experimental psychology has focused on body ownership, rather than ownership of feelings or thoughts, with several paradigms being widely used such as the rubber/virtual hand illusion [118] [119], body transfer or enfacement illusion [120], and the out-of-body illusion [88]. In all these paradigms, findings typically show a body "transfer" effect, meaning that one's own body (or body part) seems to be "displaced" from its actual location, and typically towards an artificial limb/object. This effect is due to an illusion arising from sensory stimulation (tactile and visual) applied simultaneously to one's own body part and the artificial object.

In the context of robotics, several interesting questions related to SoO can be addressed through teleoperation of humanoids. In teleoperation paradigms, the robot becomes a physically embodied avatar that the human user can control through various interfaces including, in the most advanced case, full-body motion capture [121]. Jazbec and colleagues [122] have shown that when operating an android robot, some degree of body transfer towards the robot occurs, paralleling the classical illusion [26]. Jung and colleagues [123] examined the effect of body transfer to a robot and called it the "beaming" effect, while Ventre-Dominey and colleagues [124]showed that participants experienced embodiment (or rather enfacement—perceiving the face of another person as their own face) in a robot after a short "beaming" procedure when the robot moved in a correlated manner to the movements of the participant's head. In sum, this literature shows that teleoperating a robot can induce body transfer effects. Future research could explore the conditions in which the beaming effect arises and how the body transfer experience might be impacted by different forms of robot embodiment (e.g., a child-like iCub vs. an adult-like robot embodiment) and context (e.g., teleoperating the robot to perform actions in a familiar vs. unfamiliar environment).

A second domain of robotics that can inform our understanding of SoO is exoskeletons, prosthetics and physical augmentation. Here, the crucial question is whether artificial body parts are being integrated into one's body and whether sense of ownership can extend to those artificial parts [125]. Kieliba and colleagues [126] showed that participants who were trained to use an additional robotic thumb reported increased sense of embodiment for the extra digit, following five days training, and demonstrated improved motor control of the additional thumb. These researchers also found changes in hand representation at the neural level. Specifically, the biological fingers of the hand on which the extra digit was applied became less distinctive, in terms of neural representation, following training, compared to those on the hand where the thumb augmentation was not applied.

These results confirm earlier studies using the rubber hand [118] or virtual reality [119] where sense of ownership is transferred to external or virtual objects. The benefit of using robots lies in the opportunity to understand the precise mechanisms underlying a change in sense of ownership in relation to an "alien" body. Specifically, existing research has not resolved the question of whether sense of ownership emerges from bottom-up sensorimotor integration mechanisms or whether it requires internal body maps [26]. Some authors (e.g. [127]) have reported that the body transfer illusion can occur to objects that are not shaped like a body part, such as a table or cardboard box, as long as multisensory (visuo-tactile) stimulation is spatiotemporally correlated. This finding supports the view that bottom-up sensory cues are sufficient for SoO. Other authors [128], however, highlight the importance of internal body maps having shown that anatomical, spatial, postural or textural constraints need to be met for SoO to emerge [26]. Future research using robots with different shapes or motor repertoire could help to resolve this debate.

SoA has been operationalized in the psychological literature as sense of control over the sensory outcomes of one's voluntary actions [25]. This can be measured either explicitly, by asking participants to report on their subjective experience of degree of control, or implicitly by estimating the time interval between their action and the sensory outcome [129]. Typically, for self-generated voluntary actions, the action-outcome interval appears shorter, relative to involuntary or externally generated outcomes, a phenomenon called the "intentional binding" or "temporal compression" [130]. The intentional binding phenomenon can be modulated by a wide range of factors that could potentially be explored using robotic probes such as human-like movement repertoire or form of embodiment [131-133]. Figure 2 illustrates an experiment with iCub humanoid robot [134] in which intentional binding was explored in a Sense of Joint Action (SoJA) task in which human dyads typically experience joint agency. The results showed that humans experience so-called vicarious SoA [135] over robot actions but only if the robot has human-like motor repertoire and physical embodiment [132, 133].

In the context of human-robot interaction, Ciardo has reported decreased individual SoA during joint action with a robot [136]. This effect parallels a phenomenon observed in human-human studies [137] theorised as arising from a diffusion of responsibility in social contexts. Sahaï et al. [138] have shown increased SoA in interaction with a humanoid robot, compared to a non-anthropomorphic machine. These seemingly contradictory results might be explained, however, by the nature of the task, and different socio-cognitive mechanisms involved. While the tasks in which reduction of SoA has been observed were related to avoiding losses (and thus, the results might be due to diffusion of responsibility), the task which increased SoA has been observed for humanoid robots was a joint action task where participants shared a common goal with the robot, acting as a "team". In other words, the goal was not framed in terms of potential negative consequences (avoiding losses), but rather in a positive manner as performing a task together. These studies therefore support for the view that SoA is modulated by social context (amongst other factors); perhaps reflecting the social nature of the self, as we will explore further below.

4.2. Understanding the extended self

We will explore the use of robotics in the context of the experimental understanding of the extended self through the example of theory of mind (ToM).

As noted earlier, social interactions shape our sense of self. As we learn to distinguish between ourselves and others, we come to realise that those around us are also selves. As previously noted, ToM is a mechanism for inferring the mental states of others from their observed behaviour [49], it therefore presupposes that these others have mental states. In other words, it presupposes adoption of an "intentional stance"—a strategy that humans take to explain and predict the behaviour of others by referring to their mental states [139]. The intentional stance is the default strategy adopted towards other humans, as opposed to alternative stances such as the "design stance" and the "physical stance" (see [139]). In relation to robots, however, the situation is not straightforward. Humanoid robots, owing to their human-like appearance and sometimes their behaviour, can (although not always) elicit the adoption of the intentional stance to some extent, even though they lack true mental states (see [140-142]). Moreover, this tendency can be detected from brain activity [143] and is enhanced by robot behaviour that resembles that of people [144]. Adoption of the intentional stance interacts with the experience of agency during joint action [134], suggesting that attribution of mental states might be crucial for engaging in shared attention and other mechanisms of the self.

As noted earlier, SoA is modulated by social context, indeed, it can be experienced for one's own actions and for joint actions. When we perform actions with others we experience sense of agency not only over our own actions (and their sensory outcomes) but also over actions that are performed jointly as a team and over the sensory outcomes of those jointly performed actions. This has been conceptualized as the sense of joint agency (SoJA) and studied experimentally [145-147].

Fig. 2. An example experimental paradigm for studying SoA in a human-robot interaction. The photograph illustrates a joint action paradigm where the two partners (here a human participant and the iCub humanoid robot) are responsible for complementary actions [134]. The task is to judge the occurrence (the moment in time) of an auditory beep produced by a keypress of one of the partners. If SoJA is formed, then participants should show temporal compression (temporal binding) between the keypress and the auditory tone, regardless of who actually produced the keypress (themselves or the partner). In [134], participants formed SoJA with the robot when they attributed intentional agency to it.

In the recent human-robot interaction study illustrated in Figure 2 [134] participants were ready to form a sense of joint agency with a humanoid robot, as demonstrated by both subjective temporal estimates and EEG recording of brain activity, but only if they attributed intentionality to the robot. This result implies that both individual and sense of joint agency involve similar underlying cognitive mechanisms including those linked to intentional action [130].

In sum, this collection of studies suggests that the human interpersonal self is a mechanism that is sufficiently flexible to be reused and generalized for social contexts involving non-human others, such as robots and artificial agents. We seem to readily attribute to others their distinct own selves, even if these others are robots, this accords with a wider literature on human readiness to see artificial entities as social actors [148]. Robots provide the possibility to manipulate various types of embodiment, the degree to which actions are embodied, and the extent to which intentionality is attributed to the agents. Such experiments provide insights into the mechanisms underlying various phenomena of the interpersonal self that go beyond what can be achieved solely through human-human interaction studies.

5. UNDERSTANDING THE DIVERSITY OF SELVES THROUGH ROBOTICS

The human experience of the self is broad and diverse. Whilst existing work has largely considered the neurotypical cases, robotic modelling offers the potential to explore the diversity of selves and so could contribute to a better understand of human differences and, potentially, to the treatment of disorders of the self.

Research in psychopathology and psychiatry increasingly views a range of disorders as related to aspects of self. This is clearly the case, for instance, in relation to disorders of body representation (see [74] for review), where patients might experience a limb as belonging to someone else (as in “alien hand” syndrome), or a missing limb as still present (as in “phantom limb” syndrome) [149].

Depersonalization disorder is a condition in which both SoO and SoA are affected, resulting in changes in perspectivity. In depersonalisation, the individual subjective experience is no longer anchored in the body and their sense of embodiment is weakened, if not lost. This leads to symptoms such as feeling that you are watching yourself from outside, that life is a dream, loss of experience of control over one´s own movement, and to emotional detachment or physical numbness.

The temporal self becomes disordered in amnesia, and in a variety of dementias, whilst leaving the core sense of self intact [70]. The interpersonal self also presents as differently organised in some developmental disorders including autism [63, 150] as indicated by the loss of understanding of intentions, feelings, thoughts and nonverbal communication signals. Recent research has also linked depression to changes in both the minimal and extended self [151]. For example, the sense of agency may be disturbed, causing patients to experience a lack of self-efficacy (Vogel et al. 2024) this may also be related to a disturbed sense of time, such as slowdown or deceleration of the experience of time passing (Vogel et al. 2018).

As an example of this broader field, we consider schizophrenia where different subsyndromes have been increasingly linked to disorders of the self.

Psychopathology has identified three different groups of symptoms that, broadly-speaking, define multiple schizophrenic subsyndromes [1, 152]. These include psychomotor poverty (poverty of speech, flattening of affect, retardation of action), disorganization (incoherent speech, incongruity of affect), and reality distortion (hallucinations, delusions). The scientific understanding of the self allows the reconstruction of different psychopathological symptoms.

The scientific understanding of the self allows the reinterpretation of a variety of psychopathological symptoms identified with schizophrenia. For instance, a disturbance of the SoO with regard one’s own cognitive processes could explain experiences of “thought insertion”, “thought broadcasting”, and hallucinations which are no longer being experienced as self-induced internal perceptions [153]. On the other hand, a disturbance in the SoA, involving the capacity to monitor one’s own actions [154], is evident in a second cluster of patients with schizophrenia. For instance, patients with passivity syndrome may develop problems with representing their own intentions to act, while patients with experiences of alien control of their thoughts and actions have been found to be significantly less likely to make error corrections in the absence of visual feedback indicating a defect in "central monitoring" of actions [155].

In relation to the interpersonal self, a recent meta-analysis studied links between clinical symptoms in schizophrenia and ToM impairments. Main results were that difficulties in abstract thinking and conceptual disorganization were most strongly linked to ToM whereas associations of ToM with positive symptoms and emotional symptoms including depression and anxiety were comparably small [156].

Disturbances of the transtemporal unity of self in persons with schizophrenia often arise in relation to experience of passage of time [157]. For example, patients may show a disruption of the sequence of time confusing past, present, future. The narrative self is established by autobiographical memory which organises self-related memories in the context of a coherent personal history. Many patients with schizophrenia also show symptoms of “cognitive dysmetria” as indicated by difficulties in the coordination and monitoring of processes in the retrieval, processing, and expression of information [158].

The emerging field of computational psychiatry [159] uses computational neuroscience and machine learning approaches to understanding disorders including from the perspective of modelling the self. Robotics can contribute to computational psychiatry by providing integration of self subsystems, and embodiment, a developmental trajectory, and observable behaviour that can be measured with similar metrics as applied to humans [52].

Various schizophrenia subsyndromes, and their relation to self, could be explored through robotics, by constructing a cognitive architecture that matches aspects of the neurotypical case then altering the contributions different sub-systems. A general hypothesis suggested by predictive processing accounts is that differences in the processing of prediction errors within a cognitive hierarchy could underly multiple symptoms of schizophrenia and provide a unifying theory that could be tested in robotic models [160].

We discuss two illustrative studies. First, an influential model of disturbances of SoO and SoA is the comparator model (see [52, 161] and figure 3 left) which proposes a prediction error model of the ability to recognise one's owns actions. Based on a robotic model of mirror self-recognition and the emergence of the self-other distinction, Lanillos et al. [58] have criticised the comparator model as being too simplistic (see also [162]), suggesting a "double comparator" model, where predictions of sensory outcomes is combined with learning spatio-temporal contingencies. This work provides an example of how embodied modelling provides a strong test of the sufficiency of theoretical proposals. Second, in a model developed by Yamashita and Tani [163], a two-layered network composed of a sensorimotor layer and an intentional layer, implemented as a controller for a humanoid robot, showed network level perturbations at mild levels of impairment (uncompensated error signals between layers), comparable to aberrant feelings or thoughts. However, at higher levels of impairment the robot displayed changes of overt behaviour such as disorganised or stereotyped actions, comparable to the more severe deficits seen in some patients with chronic conditions.

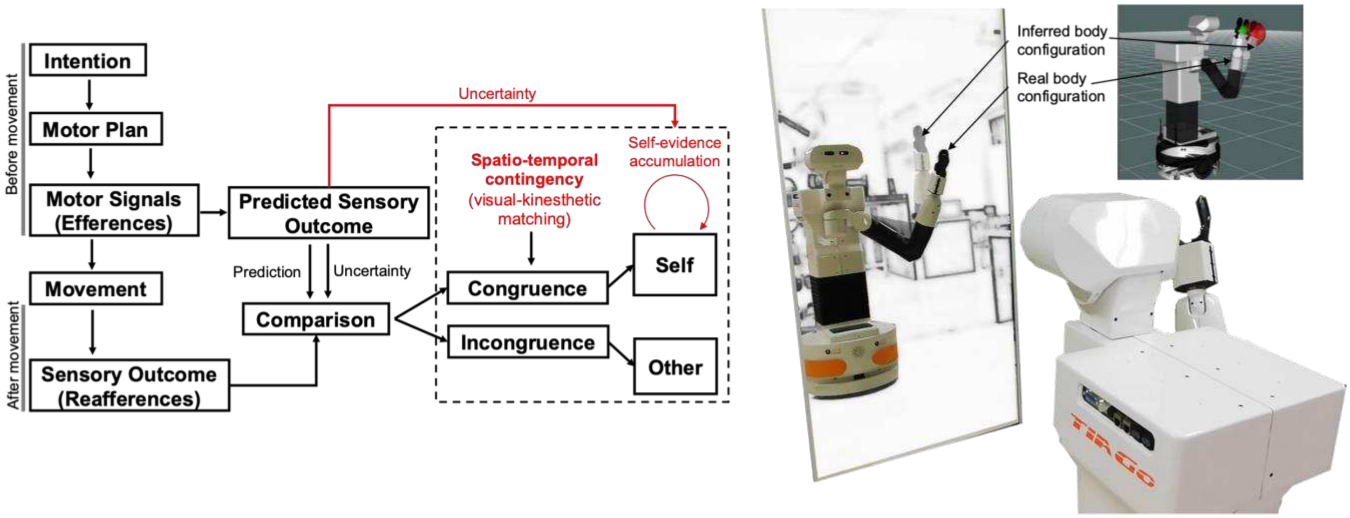

Figure 3. *Left:* A robot control system for mirror self-recognition [58] based on the comparator model [161]. Building on the reafference principle (section 2), the comparator model proposes that self-agency is detected by congruence between the predicted sensory outcomes of motor movement and observed sensory outcomes (left, black text and outlines, adapted from [161] with permission). *Right:* robot modelling of body and action recognition in a mirror [58] (permission to be obtained) suggests that the need for an additional mechanism (red text and arrows) that evaluates whether sensor events are contingent on robot's actions as previously hypothesised in [162]. Behavioural studies of people with schizophrenia indicate disturbances to these types of self-monitoring mechanisms [161] which could be better understood through robotic modelling. Adapted from [58] with permission from the authors.

From the perspective of using robots as experimental probes, robotic embodiment and manipulations that lead healthy controls to the experience of body transfer effect (e.g., enfacement) might be of interest to use to study in the disorders of the self. For example, by studying SoA in patients with psychosis using robot teleoperation, one might understand the mechanisms underlying altered SoA, altered SoO and weakened sense of self in this population [164]. Communicative accounts of psychopathology (e.g. [165]) propose the dyad, rather than the individual, as the fundamental unit of the analysis in understanding mental disorders. From this viewpoiny, robots could be used as interlocutors for patients and to simulate interactions with patients, providing a controlled environment for the systematic study of specific variants of communicative behaviour in different psychopathological conditions such as the loss of intuitive nonverbal communication capacities in autism spectrum disorder.

6. DISCUSSION

We have argued that robotics can play an important role in the scientific understanding of the self, both through the construction of embodied models of self, and through use of robots as experimental probes in paradigms that explore the human sense of self. In both cases, the embodiment of the robot, including its morphology, behaviour, appearance, and mere physical presence, allow us to develop and test increasingly refined hypotheses about the nature of self, including its development, its manifestation in behaviour, and the diversity of selves in humans, animals, and potentially machines.

Our review has largely focused on the self in cognition and action, rather than the phenomenal experience of self. This is not to exclude the study of subjectivity, as being central to understanding the self, or to suppose that this cannot ultimately be addressed through robotic studies. Indeed, our view is that constructing a robot cognitive architecture that can exhibit the capacities of the minimal and extended self, as described above, can help to operationalise notions of subjectivity such that we can ask what further might be required for an artificial entity to experience subjective states. For Blanke and Metzinger [88] the grounds for ascribing a "minimal phenomenal selfhood" require "(i) a globalized form of identification with the body as a whole (as opposed to ownership for body parts), (ii) spatiotemporal self-location and (iii) a [weak, as defined above] first-person perspective (1PP)" ([88], p. 8). On the basis of our review, we consider that research in robotics is well advanced towards achieving this degree of organisational complexity. Indeed, Gallagher [42], in his account of the first-person experience of a minimal self, discusses a robot developed by Tani [166], equipped with a predictive model, which he describes as already providing a possible instantiation of minimal phenomenal self-hood.

Other perspectives, particularly from enactivist and organismic viewpoints (e.g. [167, 168]), set the bar for subjectivity higher on the basis that robots, at least the currently existing ones, achieve only weak expression of key requirements around embodiment and agency. This critique goes beyond the limitations previously discussed, relating to the impoverished nature of robot embodiment. For instance, for Sharkey and Ziemke [167], a key distinction between current robots and animals is that the latter should be considered as "autopoietic machines" whose fundamental nature is to actively maintain their own organisation through processes including homeostasis. Biological organisms also have the character of open thermodynamic systems that actively resist their own decay. This requirement chimes with some of the broader aspects of the minimal/proto/primal self as defined by Dennett, Damasio and Panksepp. Whilst it is possible to add homeostatic and emotional mechanisms to robot control architectures (e.g. [36]) this is unlikely to satisfy stronger versions of the enactivist critique unless robots can be made genuinely self-maintaining.

We would also note that subjectivity is distinct from consciousness, although the two are not independent. Conscious states are necessary prerequisites for the full development of the self (e.g. transtemporal unity). On the other hand, conscious states are not sufficient for the self-model, as, for instance, demonstrated in meditative states [169, 170] and experiments with psychoactive drugs [171]. There is a significant literature on the possibility of creating conscious robots, readers are referred to [172] as a starting point.

There are ethical issues, in relation to creating robots with a sense of self, both with regard to status of such robots and their possible impacts on society, that require serious consideration and are worth briefly noting here (see also [148, 173, 174]). For instance, Metzinger [175] has cautioned against developing robots that have subjectivity, on the grounds that there is a risk that such entities could endure negative experiences similar to pain; in such circumstances robots could become moral patients [176]. On the other hand, advances in robotics are such that we may be nearing this point, if so, it may be best to do so knowingly and with appropriate consideration for these potential consequences [173]. Robots are increasingly being used in public life, in a variety of forms, including as socially-assistive robots that directly interact with people [148, 177]. Results show that assistive robots can decrease feelings of loneliness and anxiety, and can encourage participation in social life [148, 178], while for children, they can serve educational purposes [179] and facilitate acquisition of social and cognitive skills [180, 181]. Arguably, such robots will be more effective tools if they have aspects of a sense of self. For instance, a better sense of their own embodiment will make them safer, a sense of themselves in time will make them more effective in retrieving and applying relevant information from previous interactions, and a sense of sense others will make them more able to anticipate and meet user needs.

Conclusion

Some of the key challenges in developing advanced robots and in understanding the human self are similar. For instance, given multiple, partial, fleeting, unisensory signals about the world and the body how do we build a coherent, stable, integrated and perspectival understanding of our own embodiment and situatedness? In this review we have considered how robotics can be used to explore these questions from two fronts—through robotic modelling of the sense of self and by providing robot probes to help experimentalists explore the human sense of self. While work is still at an early stage, an emerging view is that key phenomena of self can be generated in robots with suitably configured sensor and actuator systems and a layered cognitive architecture involving networks of predictive models. Ultimately, we hope that this research will lead to an explanation for how a unified sense of self can arise in a distributed, but embodied, network of self processes and to a better understanding of the diversity of human selves.

Acknowledgments

Funding: TJP was supported in writing this article by innovate UK Funding under the UK's funding guarantee scheme for the EIC Pathfinder project CAVAA (project number 101071178). KV received funding from the Federal German Ministry of Education and Research (grant number 01GP2215). AW's contribution to this paper has been supported by the European Union under the European Innovation Council (EIC) Research and Innovation Programme, Project VALAWAI, grant agreement number 101070930. Competing interests: TJP is a director and shareholder of two robotics companies: Consequential Robotics Ltd and Bettering Our Worlds (BOW) Ltd. Neither company is expected to benefit from publication of this article. KV and AW declare that they have no competing interests.

References

1. Vogeley, K., et al., *Essential Functions of the Human Self Model Are Implemented in the Prefrontal Cortex.* Consciousness and Cognition, 1999. **8**(3): p. 343-363.
2. Kircher, T. and A.S. David, *Self-consciousness: an integrative approach from philosophy, psychopathology and the neurosciences*, in *The self in neuroscience and psychiatry.* 2003, Cambridge University Press: New York, NY, US. p. 445-473.
3. Chamberlain, C., *What Am I? Descartes's Various Ways of Considering the Self.* Journal of Modern Philosophy, 2020.
4. Hume, D., *A Treatise on Human Nature.* 1740.
5. Woźniak, M., *"I" and "Me": The Self in the Context of Consciousness.* Frontiers in psychology, 2018. **9**: p. 1656-1656.
6. Neisser, U., *Criteria for an ecological self*, in *The Self in Infancy: Theory and Research*, P. Rochat, Editor. 1995, Elsevier: Amsterdam.
7. Fivush, R. and C.A. Haden, *Autobiographical memory and the construction of the narrative self: developmental and cultural perspectives*. 2003, Mahwah, NJ: Lawrence Erlbaum Associates.

8. Dennett, D.C., *The self as a center of narrative gravity*, in *Self and Consciousness: Multiple Perspectives*, F.S. Kessel, et al., Editors. 1992, Psychology Press: Oxford. p. 103-103.
9. Bertalanffy, L.v., *General System Theory: Foundations, Development, Applications*. 1969, New York: Braziller.
10. Barandiaran, X.E., *Autonomy and Enactivism: Towards a Theory of Sensorimotor Autonomous Agency.* Topoi, 2017. **36**(3): p. 409-430.
11. Seth, A.K., *Interoceptive inference, emotion, and the embodied self.* Trends in Cognitive Sciences, 2013. **17**(11): p. 565-573.
12. Newen, A., *The Embodied Self, the Pattern Theory of Self, and the Predictive Mind.* Front Psychol, 2018. **9**: p. 2270.
13. Gallagher, S., *A Pattern Theory of Self.* Frontiers in Human Neuroscience, 2013. **7**.
14. Prescott, T.J. and D. Camilleri, *The synthetic psychology of the self*, in *Cognitive Architectures*, M.I. Aldinhas Ferreira, J. Silva Sequeira, and R. Ventura, Editors. 2019, Springer International Publishing: Cham. p. 85-104.
15. Hohwy, J., *The Sense of Self in the Phenomenology of Agency and Perception.* Psyche, 2007. **13**(1): p. 1-20.
16. Metzinger, T., *Being No One*. 2004, Cambridge, MA: MIT Press.
17. Friston, K., *Embodied inference: or "I think therefore I am, if I am what I think"*, in *The implications of embodiment: Cognition and communication.* 2011, Imprint Academic: Charlottesville, VA. p. 89-125.
18. Hohwy, J. and J. Michael, *Why should any body have a self?*, in *The Subject's Matter: Self-consciousness and the Body*, F. de Vignemont and A. Alsmith, Editors. 2017, MIT Press: Cambridge, MA.
19. Limanowski, J. and F. Blankenburg, *Minimal self-models and the free energy principle.* Frontiers in Human Neuroscience, 2013. **7**: p. 547.
20. Seth, A.K. and M. Tsakiris, *Being a Beast Machine: The Somatic Basis of Selfhood.* Trends Cogn Sci, 2018. **22**(11): p. 969-981.
21. Tani, J. and J. White, *Cognitive neurorobotics and self in the shared world, a focused review of ongoing research.* Adaptive Behavior, 2020. **30**(1): p. 81-100.
22. Kiverstein, J., *Free Energy and the Self: An Ecological–Enactive Interpretation.* Topoi, 2018. **39**(3): p. 1-16.
23. Hafner, V.V., et al., *Prerequisites for an Artificial Self.* Front Neurorobot, 2020. **14**: p. 5.
24. Georgieff, N. and M. Jeannerod, *Beyond Consciousness of External Reality: A "Who" System for Consciousness of Action and Self-Consciousness.* Consciousness and Cognition, 1998. **7**(3): p. 465-477.
25. Haggard, P., *Sense of agency in the human brain.* Nature Reviews Neuroscience, 2017. **18**(4): p. 196-207.
26. Braun, N., et al., *The Senses of Agency and Ownership: A Review.* Frontiers in Psychology, 2018. **9**.
27. Piaget, J. and B. Inhelder, *The Child's Conception of Space*. 1956, London: Routledge & Kegan Paul.
28. Nagel, T., *Brain bisection and the unity of consciousness.* Synthese, 1971. **22**(3): p. 396-413.
29. Baars, B.J., *In the theatre of consciousness: Global workspace theory, a rigorous scientific theory of consciousness.* Journal of Consciousness Studies, 1997. **4**(4): p. 292-309.
30. Bruner, J., *The Narrative Construction of Reality.* Critical Inquiry, 1991. **18**(1): p. 1-21.
31. Blumer, H., *Symbolic interactionism: Perspective and method*. 1969, Berkeley: University of California Press.

32. Chiel, H.J. and R.D. Beer, *The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment.* Trends in Neurosciences, 1997. **20**: p. 553-557.
33. Harnad, S., *Minds, machines and Searle.* Journal of Experimental & Theoretical Artificial Intelligence, 1989. **1**(1): p. 5-25.
34. Ziemke, T., *The Embodied Self: Theories, Hunches and Robot Models.* Journal of Consciousness Studies, 2007. **14**(7): p. 167-179.
35. Verschure, P.F.M.J. and T.J. Prescott, *A Living Machines approach to the sciences of mind and brain*, in *The Handbook of Living Machines: Research in Biomimetic and Biohybrid Systems*, T.J. Prescott, N. Lepora, and P.F.M.J. Verschure, Editors. 2018, OUP: Oxford, UK. p. 15-25.
36. Moulin-Frier, C., et al., *DAC-h3: A Proactive Robot Cognitive Architecture to Acquire and Express Knowledge about the World and the Self.* IEEE Transactions on Cognitive and Developmental Systems, 2018. **10**(4): p. 1005-1022.
37. Verschure, P.F., *Synthetic consciousness: the distributed adaptive control perspective.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2016. **371**(1701).
38. Wilson, S.P. and T.J. Prescott, *Scaffolding layered control architectures through constraint closure: insights into brain evolution and development.* Philos Trans R Soc Lond B Biol Sci, 2022. **377**(1844): p. 20200519.
39. Prescott, T.J. and P.F. Dominey, *Synthesizing the temporal self: robotic models of episodic and autobiographical memory.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2024. **379**(1913): p. 20230415.
40. Dennett, D.C., *The origin of selves.* Cogito, 1989. **3**: p. 163-173.
41. Jékely, G., P. Godfrey-Smith, and F. Keijzer, *Reafference and the origin of the self in early nervous system evolution.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2021. **376**(1821).
42. Gallagher, I.I., *Philosophical conceptions of the self: Implications for cognitive science.* Trends in Cognitive Sciences, 2000. **4**(1): p. 14–21.
43. Damasio, A.R., *The Feeling of What Happens: Body, Emotions and the Making of Consciousness*. 2000, London: Vintage Books.
44. Panksepp, J., *The periconscious substrates of consciousness: affective states and the evolutionary origins of the self.* Journal of Consciousness Studies, 1998. **5**(5-6): p. 566-582.
45. Rochat, P., *Self-perception and action in infancy.* Exp Brain Res, 1998. **123**: p. 102-109.
46. Rochat, P., *Self-Unity as Ground Zero of Learning and Development.* Frontiers in Psychology, 2019. **10**: p. 414.
47. Moore, C. and K. Lemmon, *The Self in Time: Developmental Perspectives*. 2001, Mahwah, NJ: Lawrence Erlbaum Associates.
48. Vasilyeva, M. and S.F. Lourenco, *Development of spatial cognition.* WIREs Cognitive Science, 2012. **3**(3): p. 349-362.
49. Baron-Cohen, S., *Mindblindness. An essay on autism and theory of mind.* . 1995, Cambridge, Massachusetts: The MIT Press.
50. Prescott, T.J., *Robot Self.* Encyclopedia of Robotics, 2021: p. 1-9.
51. Prescott, T.J. and S.P. Wilson, *Understanding brain functional architecture through robotics.* Science Robotics, 2023. **8**(78): p. eadg6014.
52. Moller, T.J., et al., *Computational models of the "active self" and its disturbances in schizophrenia.* Consciousness and Cognition, 2021. **93**(May): p. 103155.
53. Howe, M.L. and M.L. Courage, *The emergence and early development of autobiographical memory.* Psychological Review, 1997. **104**(3): p. 499-523.

54. Bongard, J., V. Zykov, and H. Lipson, *Resilient machines through continuous self-modeling.* Science, 2006. **314**(5802): p. 1118-1121.
55. Asano, Y., K. Okada, and M. Inaba, *Design principles of a human mimetic humanoid: Humanoid platform to study human intelligence and internal body system.* Science Robotics, 2017. **2**(13): p. eaaq0899.
56. Koga, Y., et al., *Self-Body Image Acquisition and Posture Generation With Redundancy Using Musculoskeletal Humanoid Shoulder Complex for Object Manipulation.* IEEE Robotics and Automation Letters, 2021. **6**(4): p. 6686-6692.
57. Schillaci, G., et al. *Is that me? Sensorimotor learning and self-other distinction in robotics*. in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 2013.
58. Lanillos, P., J. Pages, and G. Cheng, *Robot self/other distinction: Active inference meets neural networks learning in a mirror*, in *Frontiers in Artificial Intelligence and Applications*. 2020. p. 2410-2416.
59. Bechtle, S., G. Schillaci, and V.V. Hafner. *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)*. 2016.
60. Vinanzi, S., et al., *Would a robot trust you? Developmental robotics model of trust and theory of mind.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2019. **374**(1771): p. 20180032.
61. Brooks, R.A., et al. *The Cog Project: Building a Humanoid Robot*. in *Computation for Metaphors, Analogy, and Agents*. 1999. Berlin, Heidelberg: Springer Berlin Heidelberg.
62. Scassellati, B. *Imitation and Mechanisms of Joint Attention: A Developmental Structure for Building Social Skills on a Humanoid Robot*. in *Computation for Metaphors, Analogy, and Agents*. 1999. Berlin, Heidelberg: Springer Berlin Heidelberg.
63. Scassellati, B., *Theory of Mind for a Humanoid Robot.* Autonomous Robots, 2002. **12**(1): p. 13-24.
64. Metta, G., et al., *The iCub humanoid robot: An open-systems platform for research in cognitive development.* Neural Networks, 2010. **23**(8): p. 1125-1134.
65. Saegusa, R., et al. *Active motor babbling for sensorimotor learning*. in *2008 IEEE International Conference on Robotics and Biomimetics*. 2009.
66. Hoffmann, M., et al., *Robotic Homunculus: Learning of Artificial Skin Representation in a Humanoid Robot Motivated by Primary Somatosensory Cortex.* IEEE Transactions on Cognitive and Developmental Systems, 2018. **10**(2): p. 163-176.
67. Roncone, A., et al., *Peripersonal Space and Margin of Safety around the Body: Learning Visuo-Tactile Associations in a Humanoid Robot with Artificial Skin.* PLOS ONE, 2016. **11**(10): p. e0163713.
68. Fischer, T. and Y. Demiris, *Computational Modeling of Embodied Visual Perspective Taking.* IEEE Transactions on Cognitive and Developmental Systems, 2020. **12**(4): p. 723-732.
69. Pointeau, G. and P.F. Dominey, *The Role of Autobiographical Memory in the Development of a Robot Self.* Front Neurorobot, 2017. **11**: p. 27.
70. Prescott, T.J., et al., *Memory and mental time travel in humans and social robots.* Philos Trans R Soc Lond B Biol Sci, 2019. **374**(1771): p. 20180025.
71. Trevarthen, C. and J. Delafield-Butt, *Development of human consciousness*, in *The Cambridge Encyclopedia of Child Development*, E. Geangu, B. Hopkins, and S. Linkenauger, Editors. 2017, Cambridge University Press: Cambridge. p. 821-835.
72. Jeannerod, M., *The mechanism of self-recognition in humans.* Behavioural Brain Research, 2003. **142**(1): p. 1-15.

73. Koreki, A., et al., *The relationship between interoception and agency and its modulation by heartbeats: an exploratory study.* Scientific Reports, 2022. **12**(1): p. 13624.
74. Hoffmann, M., et al., *Body Schema in Robotics: A Review.* IEEE Transactions on Autonomous Mental Development, 2010. **2**(4): p. 304-324.
75. Gama, F., et al. *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)*. 2020.
76. Lanillos, P., E. Dean-Leon, and G. Cheng, *Yielding Self-Perception in Robots Through Sensorimotor Contingencies.* IEEE Transactions on Cognitive and Developmental Systems, 2017. **9**(2): p. 100-112.
77. Marcel, V., J.K. O'Regan, and M. Hoffmann. *Learning to reach to own body from spontaneous self-touch using a generative model*. in *2022 IEEE International Conference on Development and Learning (ICDL)*. 2022.
78. Laflaquière, A., et al., *Learning agent's spatial configuration from sensorimotor invariants.* Robotics and Autonomous Systems, 2015. **71**: p. 49-59.
79. Damasio, A.R., *Self Comes to Mind: Constructing the Conscious Brain*. 2010, London: Vintage Books.
80. Stapleton, M., *Steps to a "Properly Embodied" cognitive science.* Cognitive Systems Research, 2013. **22-23**: p. 1-11.
81. Roels, E., et al., *Self-healing sensorized soft robots.* Materials Today Electronics, 2022. **1**: p. 100003.
82. Laschi, C. and B. Mazzolai, *Bioinspired materials and approaches for soft robotics.* MRS Bulletin, 2021. **46**(4): p. 345-349.
83. Aubin, C.A., et al., *Towards enduring autonomous robots via embodied energy.* Nature, 2022. **602**(7897): p. 393-402.
84. Demirel, B., et al., *Distinguishing Self, Other, and Autonomy From Visual Feedback: A Combined Correlation and Acceleration Transfer Analysis.* Frontiers in Human Neuroscience, 2021. **15**.
85. Hart, J.W. and B. Scassellati. *A robotic model of the Ecological Self*. in *2011 11th IEEE-RAS International Conference on Humanoid Robots*. 2011.
86. Pang, G., G. Yang, and Z. Pang, *Review of Robot Skin: A Potential Enabler for Safe Collaboration, Immersive Teleoperation, and Affective Interaction of Future Collaborative Robots.* IEEE Transactions on Medical Robotics and Bionics, 2021. **3**(3): p. 681-700.
87. Pfeifer, R., M. Lungarella, and F. Iida, *The challenges ahead for bio-inspired 'soft' robotics.* Commun. ACM, 2012. **55**(11): p. 76–87.
88. Blanke, O. and T. Metzinger, *Full-body illusions and minimal phenomenal selfhood.* Trends in Cognitive Sciences, 2009. **13**(1): p. 7-13.
89. Fu, D., et al., *What Can Computational Models Learn From Human Selective Attention? A Review From an Audiovisual Unimodal and Crossmodal Perspective.* Frontiers in Integrative Neuroscience, 2020. **14**.
90. Mihaylova, L., et al., *Active Sensing for Robotics - A Survey.* 2002.
91. Novianto, R. and M. Williams. *The role of attention in robot self-awareness*. in *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*. 2009.
92. Ramsay, D.S. and S.C. Woods, *Clarifying the roles of homeostasis and allostasis in physiological regulation.* Psychological Review, 2014. **121**(2): p. 225-247.
93. Krichmar, J.L., *The neuromodulatory system: A framework for survival and adaptive behavior in a challenging world.* Adaptive Behavior, 2008. **16**(6): p. 385-399.

94. Rosado, O.G., et al., *Drive competition underlies effective allostatic orchestration.* Frontiers in Robotics and AI, 2022. **9**.
95. Jimenez-Rodriguez, A., et al., *A Framework for Resolving Motivational Conflict via Attractor Dynamics*, in *Biomimetic and Biohybrid Systems*, V. Vouloutsi, et al., Editors. 2020, Springer International Publishing: Cham. p. 192-203.
96. Prescott, T.J., et al. *Simulated Dopamine Modulation of a Neurorobotic Model of the Basal Ganglia*. Biomimetics, 2024. **9**, DOI: 10.3390/biomimetics9030139.
97. Krichmar, J., *A neurorobotic platform to test the influence of neuromodulatory signaling on anxious and curious behavior.* Frontiers in Neurorobotics, 2013. **7**.
98. Blakemore, S.-J. and C. Frith, *Self-awareness and action.* Current Opinion in Neurobiology, 2003. **13**(2): p. 219-224.
99. Anderson, S.R., et al., *Adaptive cancelation of self-generated sensory signals in a whisking robot.* IEEE Transactions on Robotics, 2010. **26**(6): p. 1065-1076.
100. Cheng, T., *Introduction: Sensing the self in world.* Analytic Philosophy, 2021. **62**(1): p. 57-60.
101. Hoffmann, M. *The role of self-touch experience in the formation of the self.* arXiv, 2017. DOI: https://doi.org/10.48550/arXiv.1712.07843.
102. Yamada, Y., et al., *An Embodied Brain Model of the Human Foetus.* Scientific Reports, 2016. **6**(1): p. 27893.
103. Dominey, P.F., et al. *Improving quality of life with a narrative companion*. in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 2017.
104. Rothfuss, J., et al., *Deep Episodic Memory: Encoding, Recalling, and Predicting Episodic Experiences for Robot Action Execution.* IEEE Robotics and Automation Letters, 2018. **3**(4): p. 4007-4014.
105. Karapanagiotidis, T., et al., *Tracking thoughts: Exploring the neural architecture of mental time travel during mind-wandering.* NeuroImage, 2017. **147**: p. 272-281.
106. Yu, F., et al., *NeuroSLAM: a brain-inspired SLAM system for 3D environments.* Biological Cybernetics, 2019. **113**(5): p. 515-545.
107. Hoffman, M.W., et al., *A probabilistic model of gaze imitation and shared attention.* Neural Networks, 2006. **19**(3): p. 299-310.
108. Breazeal, C. and B. Scassellati, *Robots that imitate humans.* Trends in cognitive sciences, 2002. **6**(11): p. 481-487.
109. Demiris, Y., l. Aziz-Zadeh, and J. Bonaiuto, *Information Processing in the Mirror Neuron System in Primates and Machines.* Neuroinformatics, 2014. **12**(1): p. 63–91.
110. Johnson, M. and Y. Demiris, *Perceptual perspective taking and action recognition.* International Journal of Advanced Robotic Systems, 2005. **2**(4): p. 32.
111. Asada, M., *Development of artificial empathy.* Neuroscience Research, 2015. **90**: p. 41-50.
112. Liu, D., M. Cong, and Y. Du, *Episodic Memory-Based Robotic Planning Under Uncertainty.* IEEE Transactions on Industrial Electronics, 2017. **64**(2): p. 1762-1772.
113. Kawamura, K., et al. *Development of a robot with a sense of self*. in *2005 International Symposium on Computational Intelligence in Robotics and Automation*. 2005.
114. Burghart, C., et al. *A cognitive architecture for a humanoid robot: a first approach*. in *5th IEEE-RAS International Conference on Humanoid Robots, 2005.* 2005.
115. Wykowska, A., *Robots as mirrors of the human mind.* Current Directions in Psychological Science, 2021. **30**(1): p. 34-40.
116. Wykowska, A., *Social robots to test flexibility of human social cognition.* Int J Soc Robot, 2020. **12**(6): p. 1203-1211.

117. Wykowska, A., T. Chaminade, and G. Cheng, *Embodied artificial agents for understanding human social cognition.* Philos Trans R Soc Lond B Biol Sci, 2016. **371**(1693).
118. Botvinick, M. and J. Cohen, *Rubber hands 'feel' touch that eyes see.* Nature, 1998. **391**(6669): p. 756.
119. Slater, M., et al., *Towards a digital body: the virtual arm illusion.* Frontiers in Human Neuroscience, 2008. **2**.
120. Tsakiris, M., *Looking for Myself: Current Multisensory Input Alters Self-Face Recognition.* PLOS ONE, 2008. **3**(12): p. e4040.
121. Darvish, K., et al., *Teleoperation of Humanoid Robots: A Survey.* Ieee Transactions on Robotics, 2023. **39**(3): p. 1706-1727.
122. Jazbec, M., et al., *Body-swapping experiment with an android robot Investigation of the relationship between agency and a sense of ownership toward a different body.* 2017 Ieee International Conference on Systems, Man, and Cybernetics (Smc), 2017: p. 1471-1476.
123. Jung, M., et al., *Social Telecommunication Experience with Full-Body Ownership Humanoid Robot.* International Journal of Social Robotics, 2022. **14**(9): p. 1951-1964.
124. Ventre-Dominey, J., et al., *Embodiment into a robot increases its acceptability.* Scientific Reports, 2019. **9**(1): p. 10083.
125. Forte, G., et al., *Exoskeletons for Mobility after Spinal Cord Injury: A Personalized Embodied Approach.* Journal of Personalized Medicine, 2022. **12**(3).
126. Kieliba, P., et al., *Robotic hand augmentation drives changes in neural body representation.* Science Robotics, 2021. **6**(54).
127. Armel, K.C. and V.S. Ramachandran, *Projecting sensations to external objects: evidence from skin conductance response.* Proc Biol Sci, 2003. **270**(1523): p. 1499-506.
128. Tsakiris, M., *My body in the brain: a neurocognitive model of body-ownership.* Neuropsychologia, 2010. **48**(3): p. 703-12.
129. Moore, J.W., *What Is the Sense of Agency and Why Does it Matter?* Front Psychol, 2016. **7**: p. 1272.
130. Haggard, P., S. Clark, and J. Kalogeras, *Voluntary action and conscious awareness.* Nat Neurosci, 2002. **5**(4): p. 382-5.
131. Zopf, R., V. Polito, and J. Moore, *Revisiting the link between body and agency: visual movement congruency enhances intentional binding but is not body-specific.* Scientific Reports, 2018. **8**(1): p. 196.
132. Roselli, C., et al., *Human-likeness and attribution of intentionality predict vicarious sense of agency over humanoid robot actions.* Scientific Reports, 2022. **12**(1): p. 13845.
133. Roselli, C., F. Ciardo, and A. Wykowska, *Intentions with actions: The role of intentionality attribution on the vicarious sense of agency in Human–Robot interaction.* Quarterly Journal of Experimental Psychology, 2021. **75**(4): p. 616-632.
134. Navare, U.P., et al., *When performing actions with robots, attribution of intentionality affects the sense of joint agency.* Science Robotics. **9**(91): p. eadj3665.
135. Strother, L., House, K. A., Obhi, S. S., *Subjective agency and awareness of shared actions.* Consciousness and Cognition, 2010. **19**: p. 12-20.
136. Ciardo, F., et al., *Attribution of intentional agency towards robots reduces one's own sense of agency.* Cognition, 2020. **194**: p. 104109.
137. Beyer, F., et al., *Beyond self-serving bias: diffusion of responsibility reduces sense of agency and outcome monitoring.* Social Cognitive and Affective Neuroscience, 2017. **12**(1): p. 138-145.
138. Sahai, A., et al., *Modulations of one's sense of agency during human-machine interactions: A behavioural study using a full humanoid robot.* Q J Exp Psychol (Hove), 2023. **76**(3): p. 606-620.

139. Dennett, D.C., *The Intentional Stance*. 1989, Cambridge, Massachusetts: The MIT Press.
140. Marchesi, S., et al., *Do we adopt the intentional stance toward humanoid robots?* Frontiers in Psychology, 2019. **10**: p. 450.
141. Thellman, S., A. Silvervarg, and T. Ziemke, *Folk-Psychological Interpretation of Human vs. Humanoid Robot Behavior: Exploring the Intentional Stance toward Robots.* Frontiers in Psychology, 2017. **8**.
142. Martini, M.C., C.A. Gonzalez, and E. Wiese, *Seeing Minds in Others – Can Agents with Robotic Appearance Have Human-Like Preferences?* PLOS ONE, 2016. **11**(1): p. e0146310.
143. Bossi, F., et al., *The human brain reveals resting state activity patterns that are predictive of biases in attitudes toward robots.* Science Robotics, 2020. **5**(46).
144. Marchesi, S., et al., *Belief in Sharing the Same Phenomenological Experience Increases the Likelihood of Adopting the Intentional Stance Toward a Humanoid Robot.* Technology, Mind, and Behavior, 2022. **3**(3).
145. Loehr, J.D., *The sense of agency in joint action: An integrative review.* Psychonomic Bulletin & Review, 2022. **29**(4): p. 1089-1117.
146. Jenkins, M., et al., *An investigation of "We" agency in co-operative joint actions.* Psychological Research, 2021. **85**(8): p. 3167-3181.
147. Sahaï, A., et al., *Action co-representation and the sense of agency during a joint Simon task: Comparing human and machine co-agents.* Consciousness and Cognition, 2019. **67**: p. 44-55.
148. Prescott, T.J. and J.M. Robillard, *Are friends electric? The benefits and risks of human-robot relationships.* iScience, 2021. **24**(1): p. 101993.
149. Melzack, R., et al., *Phantom limbs in people with congenital limb deficiency or amputation in early childhood.* Brain, 1997. **120**(9): p. 1603-1620.
150. Baron-Cohen, S., A.M. Leslie, and U. Frith, *Does the autistic child have a 'theory of mind'?* Cognition, 1985. **21**: p. 37-48.
151. Davey, C.G. and B.J. Harrison, *The self on its axis: a framework for understanding depression.* Translational Psychiatry, 2022. **12**(1): p. 23.
152. Liddle, P.F., *The symptoms of chronic schizophrenia: A re-examination of the positive-negative dichotomy.* The British Journal of Psychiatry, 1987. **151**: p. 145-151.
153. Kai, V. and C. Gabriel, *Pictorial pseudohallucinations with an "aperture effect" in a patient with quadrantanopia.* Journal of Neurology, Neurosurgery &amp; Psychiatry, 1998. **65**(2): p. 275.
154. Frith, C., A. Lawrence, and D. Weinberger, *The Role of the Prefrontal Cortex in Self-Consciousness: The Case of Auditory Hallucinations [and Discussion].* Philosophical Transactions: Biological Sciences, 1996. **351**(1346): p. 1505-1512.
155. Frith, C.D. and D.J. Done, *Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action.* Psychological Medicine, 1989. **19**(2): p. 359-363.
156. Thibaudeau, E., et al., *Disentangling the Relationships Between the Clinical Symptoms of Schizophrenia Spectrum Disorders and Theory of Mind: A Meta-analysis.* Schizophrenia Bulletin, 2023. **49**(2): p. 255-274.
157. Vogel, D.H.V., et al., *Disturbed time experience during and after psychosis.* Schizophrenia Research: Cognition, 2019. **17**: p. 100136.
158. Andreasen, N.C., et al., *Schizophrenia and cognitive dysmetria: a positron-emission tomography study of dysfunctional prefrontal-thalamic-cerebellar circuitry.* Proceedings of the National Academy of Sciences, 1996. **93**(18): p. 9985-9990.
159. Montague, P.R., et al., *Computational psychiatry.* Trends in Cognitive Sciences, 2012. **16**(1): p. 72-80.

160. Smith, R., P. Badcock, and K.J. Friston, *Recent advances in the application of predictive coding and active inference models within clinical neuroscience.* Psychiatry and Clinical Neurosciences, 2021. **75**(1): p. 3-13.
161. David, N., A. Newen, and K. Vogeley, *The "sense of agency" and its underlying cognitive and neural mechanisms.* Consciousness and Cognition, 2008. **17**(2): p. 523-534.
162. Zaadnoordijk, L., T.R. Besold, and S. Hunnius, *A match does not make a sense: on the sufficiency of the comparator model for explaining the sense of agency.* Neuroscience of Consciousness, 2019. **2019**(1): p. niz006.
163. Yamashita, Y. and J. Tani, *Spontaneous Prediction Error Generation in Schizophrenia.* PLOS ONE, 2012. **7**(5): p. e37843.
164. Garbarini, F., et al., *Abnormal Sense of Agency in Patients with Schizophrenia: Evidence from Bimanual Coupling Paradigm.* Frontiers in Behavioral Neuroscience, 2016. **10**.
165. Vogeley, K., *Communication as Fundamental Paradigm for Psychopathology*, in *The Oxford Handbook of 4E Cognition*. 2018, Oxford University Press. p. 0.
166. Tani, J., *An interpretation of the 'self' from the dynamical systems perspective: a constructivist approach.* Journal of Consciousness Studies, 1998. **5**: p. 516–542.
167. Sharkey, N.E. and T. Ziemke, *Mechanistic versus phenomenal embodiment: Can robot embodiment lead to strong AI?* Cognitive Systems Research, 2001. **2**(4): p. 251-262.
168. Di Paolo, E., *Organismically-Inspired Robotics: Homeostatic Adaptation and Teleology Beyond the Closed Sensorimotor Loop.* Dynamical systems approach to embodiment and sociality, 2003(August): p. 19-42.
169. Van Gordon, W., S. Sapthiang, and E. Shonin, *Contemplative Psychology: History, Key Assumptions, and Future Directions.* Perspect Psychol Sci, 2022. **17**(1): p. 99-107.
170. Holas, P. and J. Kaminska, *Mindfulness meditation and psychedelics: potential synergies and commonalities.* Pharmacol Rep, 2023. **75**(6): p. 1398-1409.
171. Letheby, C. and P. Gerrans, *Self unbound: ego dissolution in psychedelic experience.* Neuroscience of Consciousness, 2017. **2017**(1): p. nix016.
172. Chella, A. and R. Manzotti, *Artificial Consciousness*. 2011, Exeter, UK: Imprint Academic.
173. Prescott, T.J., *Robots are not just tools.* Connection Science, 2017. **29**(2): p. 142-149.
174. Gunkel, D., *Robot Rights*. 2018, Cambridge, MA: MIT Press.
175. Metzinger, T., *The Ego Tunnel: The Science of the Mind and the Myth of the Self.* 2009, New York: Basic Books.
176. Bryson, J.J., *Patiency is not a virtue: the design of intelligent systems and systems of ethics.* Ethics and Information Technology, 2018. **20**(1): p. 15-26.
177. Tapus, A., M.J. Mataric, and B. Scassellati, *Socially Assistive Robots.* IEEE RAM, 2007(March): p. 35-42.
178. Rabbitt, S.M., A.E. Kazdin, and B. Scassellati, *Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use.* Clinical Psychology Review, 2015. **35**: p. 35-46.
179. Belpaeme, T., et al., *Social robots for education: A review.* Sci Robot, 2018. **3**(21).
180. Scassellati, B., H. Admoni, and M. Mataric, *Robots for Use in Autism Research.* Annual Review of Biomedical Engineering, Vol 14, 2012. **14**: p. 275-294.
181. Ghiglino, D., et al., *Artificial scaffolding: Augmenting social cognition by means of robot technology.* Autism Research, 2023.