HUMANOID ROBOTS

Understanding the sense of self through robotics

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

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 to 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.

INTRODUCTION

Our lived experience of "me-here-now" is based on phenomena that are constitutive of the self, including experiences of body ownership and agency, both substantially reliant on a body-centered 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 affected in mental illness, leading to strong interest in the topic in the fields of psychiatry and psychopathology (*2*). As we will demonstrate in this article, the self is increasingly explored in cognitive robotics, with growing potential to inform these other realms of scientific endeavor.

We consider three different ways in which robotics can be made useful for the scientific study of self. First, robotics allows both the systematic construction of 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 apparatuses in experimental protocols that study properties of 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, because of their particular relevance for the topic of body ownership.

THEORIES OF SELF

Although Descartes considered that the self was undivided 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 unanalyzable substance and as an aggregate enmeshed with the body is reflected in hundreds of years of debate in philosophy and science about the human condition. For the psychologist William James, there were two sides to the self—one the subject of experience ("I"), and the other its object ("me") (*5*). Although philosophy and consciousness science have largely focused on the problem of subjectivity, psychology and, lately, cognitive neuroscience have developed a complementary understanding of the self as an object. For instance, a rich taxonomy was developed by Neisser, who separated various forms of self-knowledge into private, conceptual, temporally extended, interpersonal, and ecological (situated) 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*).

A systems perspective on the self

Contemporary views emphasize a broader "systems" approach in biology that sees organisms, including humans, as complex dynamical systems (*9*). Systems, as aggregates of interacting parts, can have emergent properties that are not present in any of their components. From a systems perspective, the self could be constituted by patterns of attractor dynamics in distributed neural networks, closely coupled with the environment via sensors and effectors allowing communication with other embodied selves. Such a view aligns with enactivist and embodied cognition views of selfhood (*10*–*12*), recognizing the dependency of the self on the body, on the wider environmental context, and on the construction of the self through interaction with others. Gallagher (*13*) has proposed a related view of the 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 subsystems that are at least partially independent of each other and that collectively give rise to the human sense of self (*14*).

In this review, we understand the self as a complex system that binds across subsystems 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 (i.e., a complex informational entity) instantiated by a cluster of predictive subsystems established within the brain and body and adapted to the human ecological niche (*15*–*21*).

[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–Cognitive Neuroscience (INM3), Research Center Juelich, Juelich, Germany. [4]Social Cognition in Human-Robot Interaction Unit, Italian Institute of Technology, Genova, Italy.
*Corresponding author. Email: agnieszka.wykowska@iit.it

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 (*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 grounded in a spatial, body-centered perspective (*16*, *19*, *27*).

Beyond the persistence of the body, and of our experiences as enduring in time (*28*), the transtemporal unity of the self involves the construction 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" (*29*). Moreover, this allows one to act as an "observer, agent, and guardian of the continuity of experience" [(*30*), p. 161].

As we will explore below, we consider that robotics, allied with appropriate artificial intelligence methods such as generative modeling, can provide a useful test bed to investigate this concept of the self as both a cluster of dissociable subsystems and as a constructed whole. By assembling appropriate modules within the cognitive architecture of a robot, a sense of self can emerge as contained within the boundary of the body and 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 (*31*).

Although we primarily consider key components of the self from a psychological perspective in this article, this is not to dismiss the importance of understanding how the self becomes realized physically, neurobiologically, 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, which ground cognitive capacities and the sense of self [see also (*34*)]. More broadly, we consider that a full theory of the self will encompass multiple levels of explanation (*35*) and that robotics research on the sense of self can be usefully informed by neurobiology, for instance, with regard to its realization through layered control architectures (*31*, *36*–*38*). We note that some research on the self asserts that a sense of self understood as subjective experience cannot be realized in nonbiological substrates. We will return to this important question in the Discussion.

Minimal and extended selves

From an evolutionary perspective, biological organisms distinguish processes that occur within the body from external processes. This provides the basis for a distinction between self and other, which is presumably one of the most foundational aspects of the self (*39*). 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 (*40*).

Broadly, two aspects of self—sense of body ownership (SoO) and sense of agency (SoA)—building on this primary distinction between self and other can be regarded as the central elements for the minimal self (*39*, *41*). This idea also largely coincides with Neisser's notion of the ecological self (*6*) (p. 20). Related proposals have been made by Damasio (*42*), 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 (*43*), who discusses the "primal self," emphasizing the convergence of emotional (evaluative), sensory, and motor schemata with an integrative body map. There is broad agreement that this elementary form of self needs no capacity to reflect on itself, form attitudes or beliefs about itself, or conceive or be aware of itself as persisting in time (*41*).

Insights from developmental psychology suggest the presence of at least a minimal self, and probably more, in the human infant (*44*, *45*). Over the course of development, the child will construct further aspects of the self, including the capacity for mental time travel into the past or future, that provides for a sense of self extended in time (*46*), and the awareness of itself as located in space (*47*). The child will also achieve "theory of mind" (ToM) (*48*)—the ability to infer goals, beliefs, desires, emotions, and intentions from the behavior of others and to recognize that these are different from one's own. With ToM, the child gains an understanding of social others as "other selves," building on simpler forms of awareness of others that may be present at birth or emerge during the earliest years (as indicated by joint attention, for example). In later childhood, autobiographical memory, the ability to recall and structure past episodes in propositional form, leads to the construction of a sense of self as the "narrative center" (*7*, *8*) of the child's emerging "life story." We will refer to these additional aspects of self collectively as the "extended self" (*14*, *49*).

ROBOTS AS MODELS OF THE SELF

A variety of robots have been used to investigate different aspects of the self and related behavioral phenomena (Fig. 1). The methodology of using robots to model the self follows an "understanding through building" approach that capitalizes on the value of physical models in investigating complex systems (*35*). Robots provide test beds in which candidate subsystems of the self, including models of target brain processes, can be embedded as part of a wider control architecture (*50*) 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 behavioral phenomena linked to the sense of self. Through this process, robots can be used to provide answers to some foundational questions about the self, including about the nature of the sense of self, the preconditions for its emergence, and the potential diversity of selves, including the minimal case (*14*).

Theories of the self (*15*, *17*–*20*, *22*) are increasingly based on a predictive processing view that sees the construction of the self as a process of minimizing prediction errors 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 as the structure, location, and pose of the body. At higher levels, such models can encode more abstract concepts such as memories, intentions, goals, beliefs about the self, and ultimately a concept of the self. Such a view is particularly amenable to implementation and testing via robotic embodiment (*21*, *23*, *31*, *51*).

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).

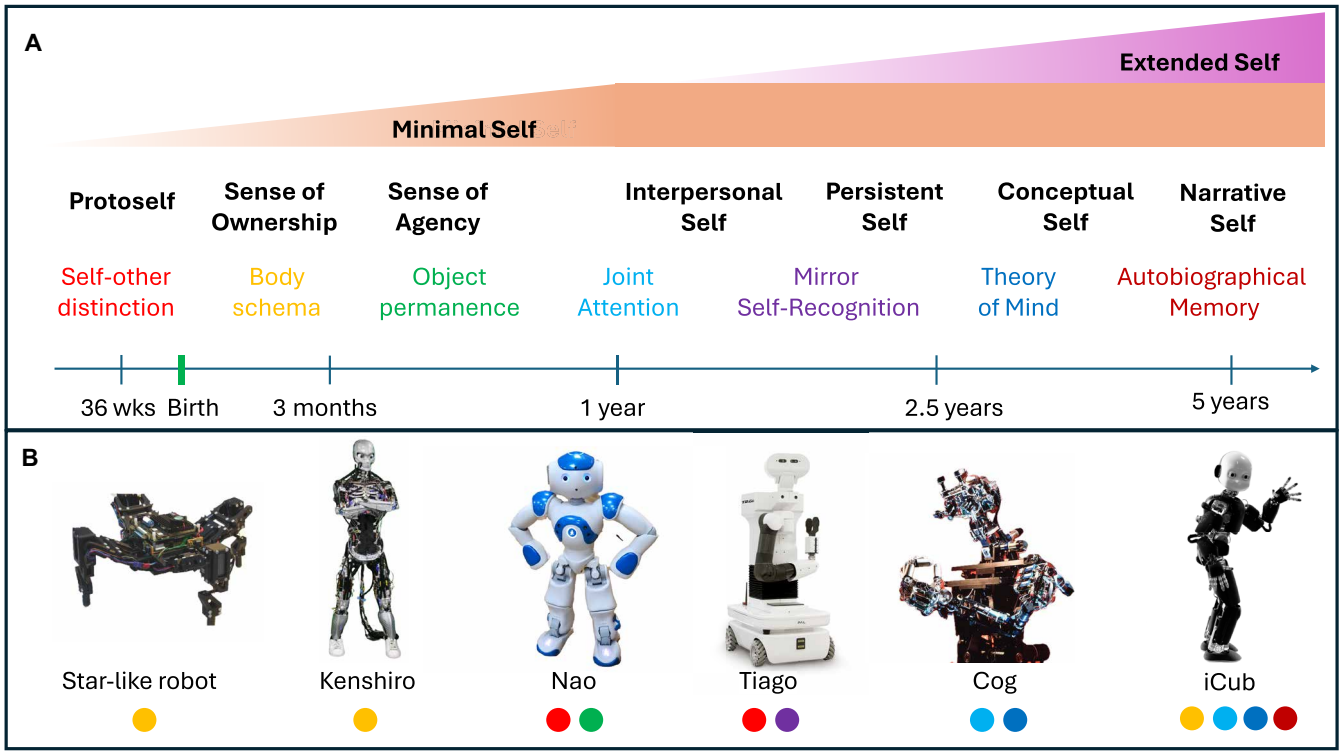

Fig. 1. The development of the human self and its investigation in robotics. (**A**) Aspects of self (black bold text) and some of the psychological capacities (colored text) associated with them, shown above an approximate developmental timeline (log scale) based on (*44*–*46*, *183*). Early-emerging aspects of self are related to the "minimal self," and later-emerging aspects are related to the "extended self." (**B**) Examples of robots used to investigate different aspects of self through embodied modeling. Colored circles indicate the corresponding psychological capacities that these robots have been programmed to emulate. From left to right: a star-like robot used to investigate the emergence of a body schema through continuous self-modeling (*55*). The Kenshiro (*184*) robot mimics the human musculoskeletal structure and is being developed as a platform for simulating the human body schema (*58*). Commercial humanoid platforms such as Nao, Pepper (not shown), and Tiago have been used to investigate phenomena associated with self-other distinction (*68*, *69*) and with the emergence of the extended self, including object permanence (*87*), ToM (*103*), and mirror-self recognition (*69*). Cog, an upper-torso humanoid, designed as a tool to study human cognitive development (*185*), was used to investigate joint attention (*99*) and ToM (*104*). The iCub humanoid (*186*) 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*, *56*, *59*), self-other distinction (*72*), joint attention (*36*), ToM (*105*), episodic/autobiographical memory, and the narrative self (*31*, *36*, *93*, *108*). These examples of embodied models of various aspects of self, coming from many different laboratories and groups, are also cited in the main text. Note that the horizontal line indicating progression from left to right relates to the developmental timeline and does not imply increasing complexity or anthropomorphism in the mentioned robot systems.

In adult humans, these capacities are reflective (i.e., accessible to introspection) and involve metacognition; however, behavioral markers indicate that infants already have rudimentary SoO and SoA but without the adult capacity to conceptualize their experience of self (*52*). In developing robot models of the minimal self, a key interest is in emulating these prereflective forms of ownership and agency.

The importance of sensing the body is broadly recognized in discussions of the self but encompasses at least two distinct themes—one focusing on proprioception and kinesthetics (*24*, *53*) and one on interoception, that is, sensing of internal bodily processes (*20*, *54*). 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, for example through self-directed exploration (*55*–*57*), that could act as a model of the human body schema (*36*, *56*–*61*). Despite a wealth of studies and data, there is still a lack of understanding of how the brain constructs and uses representations of the body. Robotics can contribute by embedding models of hypothesized 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 use actuator and mechanical systems closely modeled on human anatomy and physiology [for example, (*58*)] allowing for a similar motor repertoire. Hypotheses under examination include that latent representations that map across modalities, such as between proprioception and touch, could be useful in understanding body-related brain activity in areas such as the parietal cortex (*61*, *62*). More broadly, the capacity to learn lower-dimensional representations of the robot's sensorimotor and action spaces [for an example, see (*63*)], which discover and encode relevant invariances and establish a frame of reference with respect to the external world, 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.

Although 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 given that fundamental biological functions such as respiration, digestion, and excretion are not needed. The lack of interoception, which grounds homeostasis and emotion in

relation to these bodily processes and is often discussed as being essential for the self (*20*, *43*), has been highlighted as a potential weakness of embodied artificial systems (*64*). Research on self-healing sensorized materials for soft robots (*65*, *66*), energy harvesting (*66*), and "embodied energy" (*67*) (using power generation methods that are more similar to those used by animals) might lead to more "internally aware" robotic systems and allow the use of robots to test theories of self that integrate interoceptive mechanisms. For example, Seth and Tsakiris (*20*) have proposed that the experience of embodied selfhood arises, in part, through a process of "interoceptive inference" whereby generative models adapt to infer the causes of viscerosensory input signals. Testing such hypotheses in robotics will require the development of model systems with richer internal regulation.

The emergence of a self-other distinction has been investigated in robotic studies involving a range of modalities and mechanisms, including self-touch (*57*), visual feedback and the construction of a body model (*68*–*71*), and the construction and monitoring of peripersonal space (*72*). A key boundary of the human self is provided by the skin, which not only provides both exteroceptive and interoceptive signals (such as warmth) but also acts as a channel for both perception (e.g., shape and texture) and valence (e.g., pain and social touch). Although there has been substantial progress toward developing robotic "skin" (*73*), 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, mean that current robots have a more impoverished representation of their own embodiment than animals, although this may change with the development of soft robots and more energy-autonomous devices (*65*, *67*, *74*).

With regard to perspective taking, it is uncontroversial that a physical robot equipped with sensory systems, such as directional cameras and microphones, has a unique point of view in the literal or physical sense of there being a position in the world that is occupied by the robot and therefore cannot be occupied by anyone or anything else. There is also a specific perspective or view of the world from that position. This position, being unique to the robot at any moment, and dependent on its sensor configuration combined with body representation (as discussed above), is what Blanke and Metzinger describe as a "weak first-person perspective" (weak 1pp), requiring "a spatial frame of reference, plus a global body representation, with a perspective originating within this body representation" [(*75*), p. 7]. A weak 1pp is required for subjective experience, according to these authors, but is not, in itself, sufficient for it. Thus, a suitably designed robot with something akin to the self-manifold, as discussed above, would qualify as having a 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).

SoA is the experience of control and authorship over one's own actions and their consequences (*53*). SoA is also related to the notion of autonomy and of having a sense of choice, a minimal aspect of intentionality. An agent is autonomous to the extent that its actions are intrinsically determined rather than externally triggered and experiences choice to the extent that it perceives that alternative courses of action are possible and that its actions have observable effects.

In robotics, 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, robots can meet the core requirements (*43*, *76*) for grounding SoA in the integration of external 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 the sense of self [e.g., (*75*)] and extensively explored in robotics (*77*–*79*). SoA in biological systems is also grounded in homeostatic and allostatic regulatory systems (*38*, *80*), a layered control architecture of motivational systems (*38*), and neuromodulatory systems that are integrated with bodily processes (*81*). Models of some of these mechanisms are now being investigated in brain-inspired cognitive architectures that are engaged and embedded in the environment via robotic bodies and steered by internal value subsystems (*36*, *82*–*85*). There is a growing resemblance here to forms of biological SoA that underpin the experience of intentionality operationalized as the capacity for action selection according to intrinsic goals and preferences.

A key prerequisite for SoA in humans is to be able to predict the sensory consequences of intrinsically generated actions and thereby distinguish those consequences from other events and experience control and authorship (*86*). Such capabilities have been developed for robots (*61*, *70*, *87*–*89*) and applied to emulate milestones in infant development such as object permanence—the understanding that objects continue to exist when they are out of sight (*87*). Although philosophy has suggested the necessity of object permanence for the development of a child's ability to conceive of themself (*90*), robotics research has shown that an SoA may be required for the discovery of object permanence (*87*). 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 Fig. 1).

The phenomena associated with the minimal self, namely, SoO and SoA, both grounded in self-other distinction, are conceptually differentiated 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 (*44*, *45*), the infant has a sense both of its own body and of its own agency by the age of 2 to 3 months, many aspects 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" [(*44*), p. 108]. Robotics offers a means of operationalizing, analyzing, and distinguishing these concepts, their interdependence, and time course. For instance, the emergence of a self-other distinction before 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 with touching the lining of the womb (*91*). Simulations of a brain/body model of the human fetus (*92*) and self-touch studies in robotics have demonstrated the possibility of learning about the layout of the body using somatosensation coupled with proprioception (*59*, *62*). This approach can help identify the kinds of body representation that may already be in place before birth.

Constructing the extended self

The human self, and likely that of other complex animals, goes substantially 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 describe as the extended self. Because of space limitations, we restrict ourselves here to a brief summary of existing research.

Capacities for localizing oneself in time, which relies on episodic memory, have been investigated in the context of cognitive architectures for humanoid robots (*93*–*95*); for review, see (*31*). These studies corroborate findings from neuroscience (*96*) that similar mechanisms can support both remembering past events and imagining future events. Localization in time is widely emphasized in theories of self; however, being localized in space is also important for the sense of self and involves similar neural substrates including the hippocampal system and the default mode network (*97*). Capacities for spatial localization are well developed in robotics in the form of simultaneous localization and mapping algorithms and can be used to inform theories of subsystems in the human brain that underlie orientation and navigation (*98*).

Aspects of the interpersonal self that have been explored in robotics include phenomena like joint attention (*36*, *99*, *100*), imitation (*101*, *102*), and ToM (*103*–*107*). Some robots have been endowed with capacities for abstraction and distillation of important facts and events from episodic memory (*94*, *95*, *108*) that can allow the robot to describe and report on itself (*36*, *109*) and could provide the foundations for a narrative self and the self concept [for review, see (*31*)]. There have also been robotic investigations of the capacity to direct attention to, and inspect and reason about, the perceptions and memories generated by these systems (*79*, *110*) that could provide the beginnings of self-reflection.

The development of a humanlike 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 (*37*), which has been embodied in the iCub humanoid (*14*, *36*) and the cognitive architecture for the Karlsruhe humanoid (*95*, *111*). Evidence from studies of the human sense of self suggest the need for a layered architecture (*38*) in which core self subsystems, implemented at the level of the brainstem and available within the first few months of life, are modulated by a hierarchy of predictive subsystems specified in the cortex of the forebrain (*19*, *31*).

ROBOTS AS EXPERIMENTAL APPARATUSES 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 a model of the self. They might not yet be autonomous or endowed with any sophisticated cognitive architecture. Instead, they can be considered as technologies that augment existing experimental methodologies and apparatuses for studying human cognition (*112*).

If the research question is related to social cognition, interaction, or communication, then robots offer greater ecological validity than classical setups such as presentation of stimuli on computer screens. At the same time, robots provide improved experimental control compared with completely naturalistic protocols where participants are tested in interaction with the environment or with other humans "in the wild" (*112*–*114*). 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 SoO and SoA as core aspects of the minimal self and on mechanisms underlying the interpersonal self as an example of research related to the extended self.

Understanding the minimal self

Broadly speaking, research on SoO in experimental psychology has focused on body ownership using the rubber/virtual hand illusion (*115*, *116*), body transfer or enfacement illusion (*117*), and the out-of-body illusion (*75*), rather than ownership of feelings or thoughts. In all of 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, typically toward 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 research 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 (*118*). Jazbec and colleagues (*119*) have shown that, when operating an android robot, some degree of body transfer toward the robot occurs, paralleling the classical body transfer illusion (*26*). Jung and colleagues (*120*) examined the effect of body transfer to a robot and called it the "beaming" effect, whereas Ventre-Dominey and colleagues (*121*) 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 manner correlated 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 under which the beaming effect arises and how the body transfer experience might be influenced by different forms of robot embodiment (e.g., a childlike iCub versus an adultlike robot embodiment) and context (e.g., teleoperating the robot to perform actions in a familiar versus unfamiliar environment).

A second domain of robotics that can inform our understanding of SoO is exoskeletons, prosthetics, and physical augmentation. Here, the crucial questions are whether artificial body parts are integrated into one's body and whether SoO extends to those artificial parts (*122*). Kieliba and colleagues (*123*) showed that participants who were trained to use an additional robotic thumb reported an increased sense of embodiment for the extra digit after 5 days of 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, after training compared with those on the hand where the thumb augmentation was not applied.

These results confirm earlier studies using the rubber hand (*115*) or virtual reality (*116*) where the 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 SoO in relation to an "alien" body. Specifically, existing research has not resolved the question of whether SoO emerges from bottom-up sensorimotor integration mechanisms or whether it requires internal body maps (*26*). Armel and Ramachandran (*124*) have reported that the body transfer illusion can occur with objects that are not shaped like a body part, such as a table or cardboard box, as long as multisensory (visuotactile) stimulation is spatiotemporally correlated. This finding supports the view that bottom-up sensory cues are sufficient for SoO. Tsakiris (*125*), on the other hand, highlights 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 a 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 the participant's action and the sensory outcome (*126*). Typically, for self-generated voluntary actions, the action-outcome interval appears shorter relative to involuntary or externally generated outcomes, a phenomenon called "intentional binding," "temporal binding," or "temporal compression" (*127*). The intentional binding phenomenon can be modulated by a wide range of factors that could potentially be explored using robots, such as a humanlike movement repertoire or form of embodiment (*128*–*131*). Figure 2 illustrates an experiment with the iCub humanoid robot (*132*) 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 (*133*) over robot actions but are more likely to do so if the robot has a humanlike motor repertoire and physical embodiment (*130*, *131*).

Fig. 2. An example experimental paradigm for studying SoJA in human-robot interaction. This image illustrates a joint action paradigm where the two partners (here a human participant and the iCub humanoid robot) are responsible for complementary actions (*132*). 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 (*132*), participants formed SoJA with the robot when they attributed intentional agency to it.

In the context of human-robot interaction, Ciardo has reported decreased individual SoA during joint action with a robot (*134*). This effect parallels a phenomenon observed in human-human studies (*135*) theorized as arising from a diffusion of responsibility in social contexts. Sahaï *et al.* (*136*) have shown increased SoA in interactions with a humanoid robot compared with interactions with a nonanthropomorphic machine. These seemingly contradictory results might be explained, however, by the nature of the task and the different socio-cognitive mechanisms involved. The tasks in which reduction in SoA has been observed were related to avoiding losses (and thus the results might be due to diffusion of responsibility), but the task in 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 the view that SoA is modulated by social context (among other factors), perhaps reflecting the social nature of the self, as we will explore further below.

Understanding the extended self

The value of robotics for understanding the extended self can be illustrated with studies on ToM. As noted earlier, social interactions shape our sense of self. As we learn to distinguish between ourselves and others, we come to realize that those around us are also selves. As previously noted, ToM is a mechanism for inferring the mental states of others from their observed behavior (*48*); it therefore presupposes that these others have mental states too. This requires adoption of the "intentional stance," a strategy that humans take to explain and predict the behavior of others by referring to their mental states (*137*). The intentional stance is the default strategy adopted toward other humans, as opposed to alternative stances such as the "design stance" and the "physical stance" [see (*137*)]. In relation to robots, however, the situation is more complex. Humanoid robots, owing to their humanlike appearance and sometimes their behavior, can (although not always) elicit the adoption of the intentional stance to some extent [see (*138*–*140*)]. Moreover, this tendency can be detected from brain activity (*141*) and is enhanced by robot behavior that resembles that of humans (*142*). Adoption of the intentional stance interacts with the SoA during joint action (*132*), suggesting that attribution of mental states can affect behaviors related to the interpersonal self such as shared attention.

As noted earlier, SoA is modulated by social context; it can be experienced for one's own actions and for joint actions. When we perform actions with others, we experience SoA not only over our own actions (and their sensory outcomes) but also over actions that are performed by our partners (*130*) and over actions performed jointly as a team. As explained above, this has been conceptualized as the SoJA and studied experimentally (*143*–*145*). In the recent human-robot interaction study illustrated in Fig. 2 (*132*), participants were ready to form a SoJA with a humanoid robot, as demonstrated by both subjective temporal estimates and electroencephalography recording of brain activity, but only if they attributed intentionality to the robot. This result implies that both individual SoA and SoJA involve similar underlying cognitive mechanisms including those linked to intentional action (*127*).

In sum, this collection of studies suggests that the human interpersonal self is based on a mechanism that is sufficiently flexible to be generalized to interactions involving nonhuman others, such as robots and artificial agents. We seem to readily attribute to others their own distinct selves, even if these others are robots. This accords with the wider literature on human readiness to see artificial entities as social actors (*146*). 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.

UNDERSTANDING THE DIVERSITY OF SELVES THROUGH ROBOTICS

The human experience of self is broad and diverse. Although existing work has largely considered the neurotypical case, robotic modeling offers the potential to explore the diversity of selves and so could contribute to a better understanding of individual differences and, potentially, to the diagnosis and/or treatment of disorders of the self.

The emerging field of computational psychiatry (*147*) uses computational neuroscience and machine learning approaches to understand disorders, including from the perspective of modeling the self. Robotics can contribute by providing integrated cognitive architectures that include relevant self subsystems and exhibit embodiment, a developmental trajectory, and observable behavior that can be measured with metrics similar to those applied to humans (*51*). Specifically, cognitive architectures matching the neurotypical case could be devised, tested, and compared with versions modified to simulate diverse experiences of self, as we explore further below.

Research in psychopathology and psychiatry increasingly views a range of disorders as related to aspects of self. This is the case, for instance, in relation to disorders of body representation [see (*60*) 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) (*148*).

Depersonalization disorder is a condition in which both SoO and SoA are affected, resulting in changes in perspectivity. In depersonalization, the individual's 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 the outside, loss of experience of control over one's own movement, and possibly emotional detachment or physical numbness. Depersonalization might also be related to derealization disorder, where, for instance, life can appear to be a dream.

The temporal self becomes disordered in amnesia, and in a variety of dementias, while leaving the minimal self intact (*93*). The interpersonal self also presents as differently organized in some developmental disorders, including autism (*104*, *149*), as indicated by the loss of understanding of intentions, feelings, thoughts, and nonverbal communication signals. Recent research has linked depression to changes in both the minimal and extended self (*150*). For example, SoA may be disturbed, causing patients to experience a lack of self-efficacy (*151*); this may also be related to a disturbed sense of time, such as slowdown or deceleration of the experience of time passing (*152*).

As another example, 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 (*153*). These include psychomotor poverty (poverty of speech, flattening of affect, and retardation of action), disorganization (incoherent speech and incongruity of affect), and reality distortion (hallucinations and delusions).

The scientific understanding of the self allows the reinterpretation of a variety of psychopathological symptoms identified with schizophrenia. For instance, a disturbance of SoO with regard to one's own cognitive processes could explain experiences of "thought insertion," "thought broadcasting," and hallucinations that are no longer being experienced as self-induced internal perceptions (*154*). On the other hand, a disturbance in the SoA, involving the capacity to monitor one's own actions (*155*), 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, whereas 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 (*156*).

In relation to the interpersonal self, a recent meta-analysis studied links between clinical symptoms in schizophrenia and ToM impairments. The main results indicated 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 (*157*).

Disturbances of the transtemporal unity of self in persons with schizophrenia often arise in relation to experience of the passage of time (*158*). For example, patients may show a disruption of the sequence of time, confusing past, present, and future. The narrative self is established by autobiographical memory, which organizes self-related memories in the context of a coherent personal history. Many patients with schizophrenia show disturbances of this narrative self, including symptoms of "cognitive dysmetria," as indicated by difficulties in the coordination and monitoring of processes involved in the retrieval, processing, and expression of information (*159*).

A general hypothesis suggested by predictive processing accounts is that differences in the processing of prediction errors within a cognitive hierarchy could underlie multiple symptoms of schizophrenia. This provides a unifying theory that could be tested in robotic models (*160*), and we discuss two illustrative studies. First, an influential model of disturbances of SoO and SoA is the comparator model [see (*51*, *161*) and Fig. 3A], which proposes a prediction error account of the ability to recognize one's own actions. On the basis of a robotic investigation of mirror self-recognition and the emergence of the self-other distinction, Lanillos *et al.* (*69*) have criticized the comparator model as being too simplistic [see also (*162*)], suggesting a "double comparator" model where predictions of sensory outcomes are combined with learning of spatiotemporal contingencies. This work illustrates how embodied modeling can provide 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

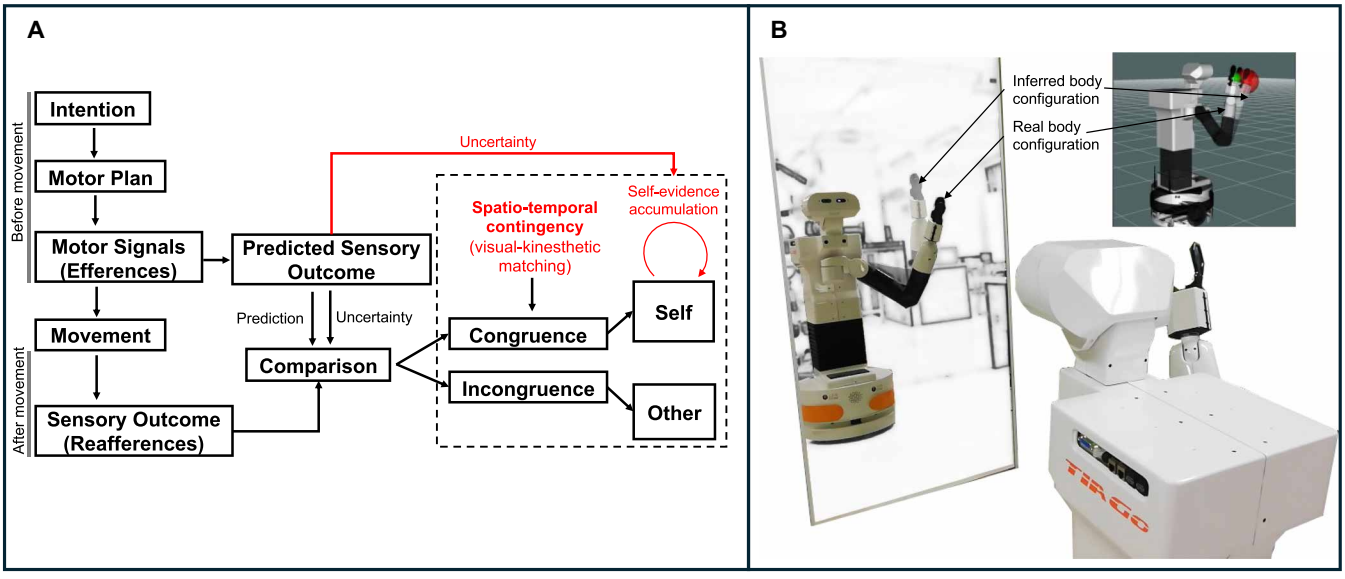

Fig. 3. A robotic investigation of the comparator model of mirror self-recognition. (**A**) A robot control system for mirror self-recognition (*69*) based on the comparator model (*161*). Building on the reafference principle ("Theories of self" section), the comparator model [black text and outlines, adapted from (*161*)] proposes that self-agency is detected by congruence between the predicted sensory outcomes of motor movement and observed sensory outcomes. (**B**) Robot modeling of body and action recognition in a mirror (*69*) suggests the need for an additional mechanism [red text and arrows in (A)] that evaluates whether sensory events are contingent on the robot's actions as previously hypothesized in (*162*). Behavioral studies of people with schizophrenia indicate disturbances to these types of self-monitoring mechanisms (*161*) that could be better understood through robotic modeling.

of impairment, the robot displayed changes of overt behavior such as disorganized or stereotyped actions, comparable to the more severe deficits seen in some patients with chronic conditions.

From the perspective of using robots as experimental tools, robotic embodiment and manipulations that lead healthy control persons to experience the body transfer effect (e.g., enfacement) might be of interest for investigations of 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 (*165*) propose the dyad, rather than the individual, as the fundamental unit of analysis in understanding mental disorders. From this viewpoint, 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 behavior under different psychopathological conditions, such as the loss of intuitive nonverbal communication capacities in autism spectrum disorder.

DISCUSSION

We have argued that robotics can play an important role in the scientific understanding of self, both through the construction of embodied models of self and through the 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, behavior, appearance, and mere physical presence, allows us to develop and test increasingly refined hypotheses about the nature of self, including its development, its manifestation in behavior, 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. 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 operationalize notions of subjectivity such that we can ask what further properties might be required for an artificial entity to experience subjective states. For Blanke and Metzinger (*75*), 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)" [(*75*), p. 8]. On the basis of our review, we consider that research in robotics is already well advanced toward achieving this degree of organizational complexity. Gallagher (*41*), 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 selfhood.

Other perspectives, particularly from enactivist and organismic viewpoints (*167*, *168*), set the bar for subjectivity higher on the basis that robots, at least 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 "autopoietic machines" whose fundamental nature is to actively maintain their own organization 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 (see the "Theories of self" section). Although it is possible to add homeostatic and emotional mechanisms to robot control architectures (*36*), this is unlikely to satisfy stronger versions of the enactivist critique unless robots can be made genuinely self-maintaining.

The possibility of realizing humanlike forms of phenomenological experience in synthetic systems rests on functionalist assumptions—that nonorganic systems could exhibit humanlike mental states (e.g., sense of self) without having to replicate the detailed physical-chemical states of human brains and bodies. Chalmers (*169*) has provided a nonreductive defense of the functionalist claim that machines could experience subjective states. We further emphasize 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 studies of meditative states (*170*, *171*) and in experiments with psychoactive drugs (*172*). There is a substantial body of literature on the possibility of creating conscious robots; readers are referred to (*173*) as a starting point.

There are ethical issues in relation to creating robots with a sense of self, both with regard to the status of such robots and their possible influence on society, that require serious consideration and are worth briefly noting here [see also (*146*, *174*, *175*)]. For instance, Metzinger (*176*) 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 (*177*). 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 (*174*). Robots are increasingly being used in public life, in a variety of forms, including as socially assistive robots that directly interact with people (*146*, *178*). Results show that assistive robots can decrease feelings of loneliness and anxiety and can encourage participation in social life (*146*, *179*); for children, they can also serve educational purposes (*180*) and facilitate acquisition of social and cognitive skills (*181*, *182*). Arguably, such robots will be more useful 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 others will make them more able to anticipate and meet human 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 approaches—through robotic modeling of the sense of self and by providing experimental probes to help researchers explore the human sense of self. Although this 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 of 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.

REFERENCES AND NOTES

1. K. Vogeley, M. Kurthen, P. Falkai, W. Maier, Essential functions of the human self model are implemented in the prefrontal cortex. *Conscious. Cogn.* **8**, 343–363 (1999).
2. T. Kircher, A. S. David, "Self-consciousness: An integrative approach from philosophy, psychopathology and the neurosciences" in *The Self in Neuroscience and Psychiatry* (Cambridge Univ. Press, 2003), pp. 445–473.
3. C. Chamberlain, What am I? Descartes's various ways of considering the self. *J. Mod. Philos.* **2**, 2147 (2020).
4. D. Hume, *A Treatise of Human Nature* (Oxford Univ. Press, 1739).
5. M. Woźniak, "I" and "Me": The self in the context of consciousness. *Front. Psychol.* **9**, 1656–1656 (2018).
6. U. Neisser, "Criteria for an ecological self" in *The Self in Infancy: Theory and Research*, P. Rochat, Ed. (Elsevier, 1995), p. 20.
7. R. Fivush, C. A. Haden, *Autobiographical Memory and the Construction of the Narrative Self: Developmental and Cultural Perspectives* (Lawrence Erlbaum Associates, 2003).
8. D. C. Dennett, "The self as a center of narrative gravity" in *Self and Consciousness: Multiple Perspectives*, F. S. Kessel, P. M. Cole, D. L. Johnson, M. D. Hakel, Eds. (Taylor & Francis, 1992), pp. 103–103.
9. L. von Bertalanffy, J. W. Sutherland, General system theory: Foundations, development, applications. *Arch. Gen. Psychiatry* **21**, 251–252 (1969).
10. X. E. Barandiaran, Autonomy and enactivism: Towards a theory of sensorimotor autonomous agency. *Topoi* **36**, 409–430 (2017).
11. A. K. Seth, Interoceptive inference, emotion, and the embodied self. *Trends Cogn. Sci.* **17**, 565–573 (2013).
12. A. Newen, The embodied self, the pattern theory of self, and the predictive mind. *Front. Psychol.* **9**, 2270 (2018).
13. S. Gallagher, A pattern theory of self. *Front. Hum. Neurosci.* **7**, 443 (2013).
14. T. J. Prescott, D. Camilleri, "The synthetic psychology of the self" in *Cognitive Architectures*, M. I. Aldinhas Ferreira, J. Silva Sequeira, R. Ventura, Eds. (Springer International Publishing, 2019), pp. 85–104.
15. J. Hohwy, The sense of self in the phenomenology of agency and perception. *Psyche* **13**, 1–20 (2007).
16. T. Metzinger, *Being No One* (MIT Press, 2003).
17. K. Friston, "Embodied inference: or 'I think therefore I am, if I am what I think'" in *The Implications of Embodiment: Cognition and Communication* (Imprint Academic, 2011), pp. 89–125.
18. J. Hohwy, J. Michael, "Why should any body have a self?" in *The Subject's Matter: Self-Consciousness and the Body*, F. de Vignemont, A. Alsmith, Eds. (MIT Press, 2017).
19. J. Limanowski, F. Blankenburg, Minimal self-models and the free energy principle. *Front. Hum. Neurosci.* **7**, 547 (2013).
20. A. K. Seth, M. Tsakiris, Being a beast machine: The somatic basis of selfhood. *Trends Cogn. Sci.* **22**, 969–981 (2018).
21. J. Tani, J. White, Cognitive neurorobotics and self in the shared world, a focused review of ongoing research. *Adapt. Behav.* **30**, 81–100 (2022).
22. J. Kiverstein, Free energy and the self: An ecological–enactive interpretation. *Topoi* **39**, 1–16 (2018).
23. V. V. Hafner, P. Loviken, A. Pico Villalpando, G. Schillaci, Prerequisites for an artificial self. *Front. Neurorobot.* **14**, 5 (2020).
24. N. Georgieff, M. Jeannerod, Beyond consciousness of external reality: A "who" system for consciousness of action and self-consciousness. *Conscious. Cogn.* **7**, 465–477 (1998).
25. P. Haggard, Sense of agency in the human brain. *Nat. Rev. Neurosci.* **18**, 196–207 (2017).
26. N. Braun, S. Debener, N. Spychala, E. Bongartz, P. Sörös, H. H. O. Müller, A. Philipsen, The senses of agency and ownership: A review. *Front. Psychol.* **9**, 535 (2018).
27. J. Piaget, B. Inhelder, *The Child's Conception of Space* (Routledge & Kegan Paul, 1956).
28. D. H. V. Vogel, M. Jording, C. Kupke, K. Vogeley, The temporality of situated cognition. *Front. Psychol.* **11**, 546212 (2020).
29. T. Nagel, Brain bisection and the unity of consciousness. *Synthese* **22**, 396–413 (1971).
30. B. J. Baars, In the theatre of consciousness: Global workspace theory, a rigorous scientific theory of consciousness. *J. Conscious. Stud.* **4**, 292–309 (1997).
31. T. J. Prescott, P. F. Dominey, Synthesising the temporal self: Robotic models of episodic and autobiographical memory. *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **379**, 20230415 (2024).

32. H. J. Chiel, R. D. Beer, The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends Neurosci.* **20**, 553–557 (1997).
33. S. Harnad, Minds, machines and Searle. *J. Exp. Theor. Artif. Intell.* **1**, 5–25 (1989).
34. T. Ziemke, The embodied self: Theories, hunches and robot models. *J. Conscious. Stud.* **14**, 167–179 (2007).
35. P. F. M. J. Verschure, 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, P. F. M. J. Verschure, Eds. (Oxford Univ. Press, 2018), pp. 15–25.
36. C. Moulin-Frier, DAC-h3: A proactive robot cognitive architecture to acquire and express knowledge about the world and the self. *IEEE Trans. Cogn. Dev. Syst.* **10**, 1005–1022 (2018).
37. P. F. Verschure, Synthetic consciousness: The distributed adaptive control perspective. *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **371**, 20150448 (2016).
38. S. P. Wilson, T. J. Prescott, Scaffolding layered control architectures through constraint closure: Insights into brain evolution and development. *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **377**, 20200519 (2022).
39. D. C. Dennett, The origin of selves. *Cogito* **3**, 163–173 (1989).
40. G. Jékely, P. Godfrey-Smith, F. Keijzer, Reafference and the origin of the self in early nervous system evolution. *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **376**, 20190764 (2021).
41. I. I. Gallagher, Philosophical conceptions of the self: Implications for cognitive science. *Trends Cogn. Sci.* **4**, 14–21 (2000).
42. A. R. Damasio, *The Feeling of What Happens: Body, Emotions and the Making of Consciousness* (Vintage Books, 2000).
43. J. Panksepp, The periconscious substrates of consciousness: Affective states and the evolutionary origins of the self. *J. Conscious. Stud.* **5**, 566–582 (1998).
44. P. Rochat, Self-perception and action in infancy. *Exp. Brain Res.* **123**, 102–109 (1998).
45. P. Rochat, Self-unity as ground zero of learning and development. *Front. Psychol.* **10**, 414 (2019).
46. C. Moore, K. Lemmon, *The Self in Time: Developmental Perspectives* (Lawrence Erlbaum Associates, 2001).
47. M. Vasilyeva, S. F. Lourenco, Development of spatial cognition. *Wiley Interdiscip. Rev. Cogn. Sci.* **3**, 349–362 (2012).
48. S. Baron-Cohen, *Mindblindness. An Essay on Autism and Theory of Mind* (MIT Press, 1995).
49. T. J. Prescott, "Robot self" in *Encyclopedia of Robotics*, M. H. Ang, O. Khatib, B. Siciliano, Eds. (Springer, 2021), pp. 1–9.
50. T. J. Prescott, S. P. Wilson, Understanding brain functional architecture through robotics. *Sci. Robot.* **8**, eadg6014 (2023).
51. T. J. Möller, Computational models of the "active self" and its disturbances in schizophrenia. *Conscious. Cogn.* **93**, 103155 (2021).
52. C. Trevarthen, J. Delafield-Butt, "Development of human consciousness" in *The Cambridge Encyclopedia of Child Development*, E. Geangu, B. Hopkins, S. Linkenauger, Eds. (Cambridge Univ. Press, 2017), pp. 821–835.
53. M. Jeannerod, The mechanism of self-recognition in humans. *Behav. Brain Res.* **142**, 1–15 (2003).
54. A. Koreki, The relationship between interoception and agency and its modulation by heartbeats: An exploratory study. *Sci. Rep.* **12**, 13624 (2022).
55. J. Bongard, V. Zykov, H. Lipson, Resilient machines through continuous self-modeling. *Science* **314**, 1118–1121 (2006).
56. R. Saegusa, G. Metta, G. Sandini, S. Sakka, Active motor babbling for sensorimotor learning, in *2008 IEEE International Conference on Robotics and Biomimetics* (IEEE, 2009), pp. 794–799.
57. F. Gama, M. Shcherban, M. Rolf, M. Hoffmann, 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)* (IEEE, 2020), pp. 1–8.
58. Y. Koga, Self-body image acquisition and posture generation with redundancy using musculoskeletal humanoid shoulder complex for object manipulation. *IEEE Robot. Autom. Lett.* **6**, 6686–6692 (2021).
59. M. Hoffmann, Robotic homunculus: Learning of artificial skin representation in a humanoid robot motivated by primary somatosensory cortex. *IEEE Trans. Cogn. Dev. Syst.* **10**, 163–176 (2018).
60. M. Hoffmann, H. Marques, A. Arieta, H. Sumioka, M. Lungarella, R. Pfeifer, Body schema in robotics: A review. *IEEE Trans. Auton. Ment. Dev.* **2**, 304–324 (2010).
61. P. Lanillos, E. Dean-Leon, G. Cheng, Yielding self-perception in robots through sensorimotor contingencies. *IEEE Trans. Cogn. Dev. Syst.* **9**, 100–112 (2017).
62. V. Marcel, J. K. O'Regan, 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)* (IEEE, 2022), pp. 328–335.
63. A. Laflaquière, J. K. O'Regan, S. Argentieri, B. Gas, A. V. Terekhov, Learning agent's spatial configuration from sensorimotor invariants. *Robot. Auton. Syst.* **71**, 49–59 (2015).
64. M. Stapleton, Steps to a "Properly Embodied" cognitive science. *Cogn. Syst. Res.* **22–23**, 1–11 (2013).
65. E. Roels, S. Terryn, J. Brancart, F. Sahraeeazartamar, F. Clemens, G. van Assche, B. Vanderborght, Self-healing sensorized soft robots. *Mater. Today Electron.* **1**, 100003 (2022).
66. C. Laschi, B. Mazzolai, Bioinspired materials and approaches for soft robotics. *MRS Bull.* **46**, 345–349 (2021).
67. C. A. Aubin, Towards enduring autonomous robots via embodied energy. *Nature* **602**, 393–402 (2022).
68. G. Schillaci, V. V. Hafner, B. Lara, M. Grosjean, Is that me? Sensorimotor learning and self-other distinction in robotics, in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (IEEE, 2013), pp. 223–224.
69. P. Lanillos, J. Pages, G. Cheng, Robot self/other distinction: Active inference meets neural networks learning in a mirror, in *24th European Conference on Artificial Intelligence, 29 August–8 September 2020, Santiago de Compostela, Spain – Including 10th Conference on Prestigious Applications of Artificial Intelligence (PAIS 2020)*, vol. 325 of *Frontiers in Artificial Intelligence and Applications* (IOS Press, 2020), pp. 2410–2416.
70. B. Demirel, C. Moulin-Frier, X. D. Arsiwalla, P. F. M. J. Verschure, M. Sánchez-Fibla, Distinguishing self, other, and autonomy from visual feedback: A combined correlation and acceleration transfer analysis. *Front. Hum. Neurosci.* **15**, 560657 (2021).
71. J. W. Hart, B. Scassellati, A robotic model of the ecological self, in *2011 11th IEEE-RAS International Conference on Humanoid Robots* (IEEE, 2011), pp. 682–688.
72. A. Roncone, Peripersonal space and margin of safety around the body: Learning visuo-tactile associations in a humanoid robot with artificial skin. *PLOS ONE* **11**, e0163713 (2016).
73. G. Pang, G. Yang, Z. Pang, Review of robot skin: A potential enabler for safe collaboration, immersive teleoperation, and affective interaction of future collaborative robots. *IEEE Trans. Med. Robot. Bionics* **3**, 681–700 (2021).
74. R. Pfeifer, M. Lungarella, F. Iida, The challenges ahead for bio-inspired 'soft' robotics. *Commun. ACM* **55**, 76–87 (2012).
75. O. Blanke, T. Metzinger, Full-body illusions and minimal phenomenal selfhood. *Trends Cogn. Sci.* **13**, 7–13 (2009).
76. A. R. Damasio, *Self Comes to Mind: Constructing the Conscious Brain* (Vintage Books, 2010).
77. D. Fu, C. Weber, G. Yang, M. Kerzel, W. Nan, P. Barros, H. Wu, X. Liu, S. Wermter, What can computational models learn from human selective attention? A review from an audiovisual unimodal and crossmodal perspective. *Front. Integr. Neurosci.* **14**, 10 (2020).
78. L. S. Mihaylova, H. Bruyninckx, J. De Scutter, Active sensing of a nonholonomic wheeled mobile robot, in *Proceedings of the 3rd IEEE Benelux Signal Processing Symposium (SPS-2002)* (IEEE, 2002), pp. 125–127.
79. R. Novianto, 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* (IEEE, 2009), pp. 1047–1053.
80. D. S. Ramsay, S. C. Woods, Clarifying the roles of homeostasis and allostasis in physiological regulation. *Psychol. Rev.* **121**, 225–247 (2014).
81. J. L. Krichmar, The neuromodulatory system: A framework for survival and adaptive behavior in a challenging world. *Adapt. Behav.* **16**, 385–399 (2008).
82. O. G. Rosado, A. F. Amil, I. T. Freire, P. F. M. J. Verschure, Drive competition underlies effective allostatic orchestration. *Front. Robot. AI* **9**, 1052998 (2022).
83. A. Jimenez-Rodriguez, T. J. Prescott, R. Schmidt, S. Wilson, "A framework for resolving motivational conflict via attractor dynamics" in *Biomimetic and Biohybrid Systems*, V. Vouloutsi, A. Mura, F. Tauber, T. Speck, T. J. Prescott, P. F. M. J. Verschure, Eds. (Springer, 2020), pp. 192–203.
84. T. J. Prescott, F. M. Montes González, K. Gurney, M. D. Humphries, P. Redgrave, Simulated dopamine modulation of a neurorobotic model of the basal ganglia. *Biomimetics* **9**, 139 (2024).
85. J. Krichmar, A neurorobotic platform to test the influence of neuromodulatory signaling on anxious and curious behavior. *Front. Neurorobot.* **7**, 1 (2013).
86. S. J. Blakemore, C. Frith, Self-awareness and action. *Curr. Opin. Neurobiol.* **13**, 219–224 (2003).
87. S. Bechtle, G. Schillaci, 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)* (IEEE, 2016), pp. 166–171.
88. S. R. Anderson, M. J. Pearson, A. Pipe, T. Prescott, P. Dean, J. Porrill, Adaptive cancelation of self-generated sensory signals in a whisking robot. *IEEE. Trans. Robot.* **26**, 1065–1076 (2010).
89. A. Stoytchev, Self-detection in robots: A method based on detecting temporal contingencies. *Robotica* **29**, 1–21 (2011).
90. T. Cheng, Introduction: Sensing the self in world. *Anal. Philos.* **62**, 57–60 (2021).
91. M. Hoffmann, The role of self-touch experience in the formation of the self. arXiv:1712.07843 [q-bio.NC] (2017).

92. Y. Yamada, An embodied brain model of the human foetus. *Sci. Rep.* **6**, 27893 (2016).
93. T. J. Prescott, Memory and mental time travel in humans and social robots. *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **374**, 20180025 (2019).
94. P. F. Dominey, V. Paléologue, A. K. Pandey, J. Ventre-Dominey, Improving quality of life with a narrative companion, in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (IEEE, 2017), pp. 127–134.
95. J. Rothfuss, F. Ferreira, E. E. Aksoy, Y. Zhou, T. Asfour, Deep episodic memory: Encoding, recalling, and predicting episodic experiences for robot action execution. *IEEE Robot. Autom. Lett.* **3**, 4007–4014 (2018).
96. D. L. Schacter, D. R. Addis, D. Hassabis, V. C. Martin, R. N. Spreng, K. K. Szpunar, The future of memory: Remembering, imagining, and the brain. *Neuron* **76**, 677–694 (2012).
97. T. Karapanagiotidis, B. C. Bernhardt, E. Jefferies, J. Smallwood, Tracking thoughts: Exploring the neural architecture of mental time travel during mind-wandering. *Neuroimage* **147**, 272–281 (2017).
98. F. Yu, J. Shang, Y. Hu, M. Milford, NeuroSLAM: A brain-inspired SLAM system for 3D environments. *Biol. Cybern.* **113**, 515–545 (2019).
99. B. Scassellati, "Imitation and mechanisms of joint attention: A developmental structure for building social skills on a humanoid robot" in *Computation for Metaphors*, *Analogy*, *and Agents* (Springer, 1999), pp. 176–195.
100. M. W. Hoffman, D. B. Grimes, A. P. Shon, R. P. N. Rao, A probabilistic model of gaze imitation and shared attention. *Neural Netw.* **19**, 299–310 (2006).
101. C. Breazeal, B. Scassellati, Robots that imitate humans. *Trends Cogn. Sci.* **6**, 481–487 (2002).
102. Y. Demiris, L. Aziz-Zadeh, J. Bonaiuto, Information processing in the mirror neuron system in primates and machines. *Neuroinformatics* **12**, 63–91 (2014).
103. S. Vinanzi, Would a robot trust you? Developmental robotics model of trust and theory of mind. *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **374**, 20180032 (2019).
104. B. Scassellati, Theory of mind for a humanoid robot. *Auton. Robots* **12**, 13–24 (2002).
105. T. Fischer, Y. Demiris, Computational modeling of embodied visual perspective taking. *IEEE Trans. Cogn. Dev. Syst.* **12**, 723–732 (2020).
106. M. Johnson, Y. Demiris, Perceptual perspective taking and action recognition. *Int. J. Adv. Robot. Syst.* **2**, 32 (2005).
107. M. Asada, Development of artificial empathy. *Neurosci. Res.* **90**, 41–50 (2015).
108. G. Pointeau, P. F. Dominey, The role of autobiographical memory in the development of a robot self. *Front. Neurorobot.* **11**, 27 (2017).
109. D. Liu, M. Cong, Y. Du, Episodic memory-based robotic planning under uncertainty. *IEEE Trans. Ind. Electron.* **64**, 1762–1772 (2017).
110. K. Kawamura, W. Dodd, P. Ratanaswasd, R. A. Gutierrez, Development of a robot with a sense of self, in *2005 International Symposium on Computational Intelligence in Robotics and Automation* (IEEE, 2005), pp. 211–217.
111. C. Burghart, R. Mikut, R. Stiefelhagen, T. Asfour, H. Holzapfel, P. Steinhaus, R. Dillmann, A cognitive architecture for a humanoid robot: A first approach, in *5th IEEE-RAS International Conference on Humanoid Robots*, *2005* (IEEE, 2005), pp. 357–362.
112. A. Wykowska, Robots as mirrors of the human mind. *Curr. Dir. Psychol. Sci.* **30**, 34–40 (2021).
113. A. Wykowska, Social robots to test flexibility of human social cognition. *Int. J. Soc. Robot.* **12**, 1203–1211 (2020).
114. A. Wykowska, T. Chaminade, G. Cheng, Embodied artificial agents for understanding human social cognition. *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **371**, 375 (2016).
115. M. Botvinick, J. Cohen, Rubber hands 'feel' touch that eyes see. *Nature* **391**, 756 (1998).
116. M. Slater, D. Perez-Marcos, H. H. Ehrsson, M. V. Sanchez-Vives, Towards a digital body: The virtual arm illusion. *Front. Hum. Neurosci.* **2**, 181 (2008).
117. M. Tsakiris, Looking for myself: Current multisensory input alters self-face recognition. *PLOS ONE* **3**, e4040 (2008).
118. K. Darvish, L. Penco, J. Ramos, R. Cisneros, J. Pratt, E. Yoshida, S. Ivaldi, D. Pucci, Teleoperation of humanoid robots: A survey. *IEEE Trans. Robot.* **39**, 1706–1727 (2023).
119. M. Jazbec, S. Nishio, H. Ishiguro, H. Kuzuoka, M. Okubo, C. Peñaloza, Body-swapping experiment with an android robot Investigation of the relationship between agency and a sense of ownership toward a different body, in *2017 IEEE International Conference on Systems*, *Man*, *and Cybernetics (SMC)* (IEEE, 2017), pp. 1471–1476.
120. M. Jung, J. Kim, K. Han, K. Kim, Social telecommunication experience with full-body ownership humanoid robot. *Int. J. Soc. Robot.* **14**, 1951–1964 (2022).
121. J. Ventre-Dominey, G. Gilbert, M. Bosse-Platiere, A. Farne, P. F. Dominey, F. Pavani, Embodiment into a robot increases its acceptability. *Sci. Rep.* **9**, 10083 (2019).
122. G. Forte, E. Leemhuis, F. Favieri, M. Casagrande, A. M. Giannini, L. de Gennaro, M. Pazzaglia, Exoskeletons for mobility after spinal cord injury: A personalized embodied approach. *J. Pers. Med.* **12**, 380 (2022).
123. P. Kieliba, D. Clode, R. O. Maimon-Mor, T. R. Makin, Robotic hand augmentation drives changes in neural body representation. *Sci. Robot.* **6**, eabd7935 (2021).
124. K. C. Armel, V. S. Ramachandran, Projecting sensations to external objects: Evidence from skin conductance response. *Proc. Biol. Sci.* **270**, 1499–1506 (2003).
125. M. Tsakiris, My body in the brain: A neurocognitive model of body-ownership. *Neuropsychologia* **48**, 703–712 (2010).
126. J. W. Moore, What is the sense of agency and why does it matter? *Front. Psychol.* **7**, 1272 (2016).
127. P. Haggard, S. Clark, J. Kalogeras, Voluntary action and conscious awareness. *Nat. Neurosci.* **5**, 382–385 (2002).
128. D. H. V. Vogel, M. Jording, C. Esser, P. H. Weiss, K. Vogeley, Temporal binding is enhanced in social contexts. *Psychon. Bull. Rev.* **28**, 1545–1555 (2021).
129. R. Zopf, V. Polito, J. Moore, Revisiting the link between body and agency: Visual movement congruency enhances intentional binding but is not body-specific. *Sci. Rep.* **8**, 196 (2018).
130. C. Roselli, F. Ciardo, D. de Tommaso, A. Wykowska, Human-likeness and attribution of intentionality predict vicarious sense of agency over humanoid robot actions. *Sci. Rep.* **12**, 13845 (2022).
131. C. F. Roselli, A. Wykowska, Intentions with actions: The role of intentionality attribution on the vicarious sense of agency in human–robot interaction. *Q. J. Exp. Psychol.* **75**, 616–632 (2021).
132. U. P. Navare, F. Ciardo, K. Kompatsiari, D. De Tommaso, A. Wykowska, Performing actions with robots: Attribution of intentionality affects the sense of joint agency. *Sci. Robot.* **9**, eadj3665 (2024).
133. L. Strother, K. A. House, S. S. Obhi, Subjective agency and awareness of shared actions. *Conscious. Cogn.* **19**, 12–20 (2010).
134. F. Ciardo, Attribution of intentional agency towards robots reduces one's own sense of agency. *Cognition* **194**, 104109 (2020).
135. F. Beyer, Beyond self-serving bias: Diffusion of responsibility reduces sense of agency and outcome monitoring. *Soc. Cogn. Affect. Neurosci.* **12**, 138–145 (2017).
136. A. Sahaï, E. Caspar, A. De Beir, O. Grynszpan, E. Pacherie, B. Berberian, Modulations of one's sense of agency during human-machine interactions: A behavioural study using a full humanoid robot. *Q. J. Exp. Psychol. (Hove)* **76**, 606–620 (2023).
137. D. C. Dennett, *The Intentional Stance* (MIT Press, 1988).
138. S. Marchesi, D. Ghiglino, F. Ciardo, J. Perez-Osorio, E. Baykara, A. Wykowska, Do we adopt the intentional stance toward humanoid robots? *Front. Psychol.* **10**, 450 (2019).
139. S. Thellman, A. Silvervarg, T. Ziemke, Folk-psychological interpretation of human vs humanoid robot behavior: Exploring the intentional stance toward robots. *Front. Psychol.* **8**, 1962 (2017).
140. M. C. Martini, C. A. Gonzalez, E. Wiese, Seeing minds in others–Can agents with robotic appearance have human-like preferences? *PLOS ONE* **11**, e0146310 (2016).
141. F. Bossi, C. Willemse, J. Cavazza, S. Marchesi, V. Murino, A. Wykowska, The human brain reveals resting state activity patterns that are predictive of biases in attitudes toward robots. *Sci. Robot.* **5**, eabb6652 (2020).
142. S. Marchesi, D. De Tommaso, J. Perez-Osorio, A. Wykowska, Belief in sharing the same phenomenological experience increases the likelihood of adopting the intentional stance toward a humanoid robot. *Technol. Mind Behav.* **3**, 1–11 (2022).
143. J. D. Loehr, The sense of agency in joint action: An integrative review. *Psychon. Bull. Rev.* **29**, 1089–1117 (2022).
144. M. Jenkins, O. Esemezie, V. Lee, M. Mensingh, K. Nagales, S. S. Obhi, An investigation of "We" agency in co-operative joint actions. *Psychol. Res.* **85**, 3167–3181 (2021).
145. A. Sahaï, A. Desantis, O. Grynszpan, E. Pacherie, B. Berberian, Action co-representation and the sense of agency during a joint Simon task: Comparing human and machine co-agents. *Conscious. Cogn.* **67**, 44–55 (2019).
146. T. J. Prescott, J. M. Robillard, Are friends electric? The benefits and risks of human-robot relationships. *iScience* **24**, 101993 (2021).
147. P. R. Montague, R. J. Dolan, K. J. Friston, P. Dayan, Computational psychiatry. *Trends Cogn. Sci.* **16**, 72–80 (2012).
148. R. Melzack, R. Israel, R. Lacroix, G. Schultz, Phantom limbs in people with congenital limb deficiency or amputation in early childhood. *Brain* **120**, 1603–1620 (1997).
149. S. Baron-Cohen, A. M. Leslie, U. Frith, Does the autistic child have a 'theory of mind'? *Cognition* **21**, 37–46 (1985).
150. C. G. Davey, B. J. Harrison, The self on its axis: A framework for understanding depression. *Transl. Psychiatry* **12**, 23 (2022).
151. D. H. V. Vogel, M. Jording, P. H. Weiss, K. Vogeley, Temporal binding and sense of agency in major depression. *Front. Psychiatry* **15**, 1288674 (2024).
152. D. H. V. Vogel, K. Krämer, T. Schoofs, C. Kupke, K. Vogeley, Disturbed experience of time in depression—Evidence from content analysis. *Front. Hum. Neurosci.* **12**, 66 (2018).
153. P. F. Liddle, The symptoms of chronic schizophrenia: A re-examination of the positive-negative dichotomy. *Br. J. Psychiatry* **151**, 145–151 (1987).
154. K. Vogeley, Hallucinations emerge from a disbalance of self and reality monitoring. *Monist* **82**, 626–644 (1999).
155. C. Frith, A. Lawrence, D. Weinberger, The role of the prefrontal cortex in self-consciousness: The case of auditory hallucinations. *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **351**, 1501–1512 (1996).
156. C. D. Frith, D. J. Done, Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychol. Med.* **19**, 359–363 (1989).

157. E. Thibaudeau, J. Rae, D. Raucher-Chéné, A. Bougeard, M. Lepage, Disentangling the relationships between the clinical symptoms of schizophrenia spectrum disorders and theory of mind: A meta-analysis. *Schizophr. Bull.* **49**, 255–274 (2023).
158. D. H. V. Vogel, Disturbed time experience during and after psychosis. *Schizophr. Res. Cogn.* **17**, 100136 (2019).
159. N. C. Andreasen, D. S. O'Leary, T. Cizadlo, S. Arndt, K. Rezai, L. L. Ponto, G. L. Watkins, R. D. Hichwa, Schizophrenia and cognitive dysmetria: A positron-emission tomography study of dysfunctional prefrontal-thalamic-cerebellar circuitry. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9985–9990 (1996).
160. R. Smith, P. Badcock, K. J. Friston, Recent advances in the application of predictive coding and active inference models within clinical neuroscience. *Psychiatry Clin. Neurosci.* **75**, 3–13 (2021).
161. N. David, A. Newen, K. Vogeley, The "sense of agency" and its underlying cognitive and neural mechanisms. *Conscious. Cogn.* **17**, 523–534 (2008).
162. L. Zaadnoordijk, T. R. Besold, S. Hunnius, A match does not make a sense: On the sufficiency of the comparator model for explaining the sense of agency. *Neurosci. Conscious.* **2019**, niz006 (2019).
163. Y. Yamashita, J. Tani, Spontaneous prediction error generation in schizophrenia. *PLOS ONE* **7**, e37843 (2012).
164. F. Garbarini, A. Mastropasqua, M. Sigaudo, M. Rabuffetti, A. Piedimonte, L. Pia, P. Rocca, Abnormal sense of agency in patients with schizophrenia: Evidence from bimanual coupling paradigm. *Front. Behav. Neurosci.* **10**, 43 (2016).
165. K. Vogeley, "Communication as fundamental paradigm for psychopathology" in *The Oxford Handbook of 4E Cognition* (Oxford Univ. Press, 2018).
166. J. Tani, An interpretation of the 'self' from the dynamical systems perspective: A constructivist approach. *J. Conscious. Stud.* **5**, 516–542 (1998).
167. N. E. Sharkey, T. Ziemke, Mechanistic versus phenomenal embodiment: Can robot embodiment lead to strong AI? *Cog. Syst. Res.* **2**, 251–262 (2001).
168. E. Di Paolo, "Organismically-inspired robotics: Homeostatic adaptation and teleology beyond the closed sensorimotor loop" in *Dynamical Systems Approach to Embodiment and Sociality*, K. Murase, T. Asakura, Eds. (Advanced Knowledge International, 2003), pp. 19–42.
169. D. J. Chalmers, A computational foundation for the study of cognition. *J. Cogn. Sci.* **12**, 323–357 (2011).
170. W. Van Gordon, S. Sapthiang, E. Shonin, Contemplative psychology: History, key assumptions, and future directions. *Perspect. Psychol. Sci.* **17**, 99–107 (2022).
171. P. Holas, J. Kaminska, Mindfulness meditation and psychedelics: Potential synergies and commonalities. *Pharmacol. Rep.* **75**, 1398–1409 (2023).
172. C. Letheby, P. Gerrans, Self unbound: Ego dissolution in psychedelic experience. *Neurosci. Conscious.* **2017**, nix016 (2017).
173. A. Chella, R. Manzotti, *Artificial Consciousness* (Imprint Academic, 2011).
174. T. J. Prescott, Robots are not just tools. *Connect. Sci.* **29**, 142–149 (2017).
175. D. Gunkel, *Robot Rights* (MIT Press, 2018).
176. T. Metzinger, *The Ego Tunnel: The Science of the Mind and the Myth of the Self* (Basic Books, 2009).
177. J. J. Bryson, Patiency is not a virtue: The design of intelligent systems and systems of ethics. *Ethics Inf. Technol.* **20**, 15–26 (2018).
178. A. Tapus, M. J. Mataric, B. Scassellati, Socially assistive robots [grand challenges of robotics]. *IEEE Robot. Autom. Mag.* **14**, 35–42 (2007).
179. S. M. Rabbitt, A. E. Kazdin, B. Scassellati, Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use. *Clin. Psychol. Rev.* **35**, 35–46 (2015).
180. T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, F. Tanaka, Social robots for education: A review. *Sci. Robot.* **3**, eaat5954 (2018).
181. B. Scassellati, H. Admoni, M. Mataric, Robots for use in autism research. *Annu. Rev. Biomed. Eng.* **14**, 275–294 (2012).
182. D. Ghiglino, F. Floris, D. de Tommaso, K. Kompatsiari, P. Chevalier, T. Priolo, A. Wykowska, Artificial scaffolding: Augmenting social cognition by means of robot technology. *Autism Res.* **16**, 997–1008 (2023).
183. M. L. Howe, M. L. Courage, The emergence and early development of autobiographical memory. *Psychol. Rev.* **104**, 499–523 (1997).
184. Y. Asano, K. Okada, M. Inaba, Design principles of a human mimetic humanoid: Humanoid platform to study human intelligence and internal body system. *Sci. Robot.* **2**, eaaq0899 (2017).
185. R. A. Brooks, *The Cog Project: Building a Humanoid Robot*. in *Computation for Metaphors, Analogy, and Agents* (Springer, 1999).
186. G. Metta, The iCub humanoid robot: An open-systems platform for research in cognitive development. *Neural Netw.* **23**, 1125–1134 (2010).

Acknowledgments

Funding: T.J.P. 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). K.V. received funding from the Federal German Ministry of Education and Research (grant number 01GP2215). A.W.'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. **Author contributions:** Conceptualization: T.J.P., K.V., and A.W. Writing—original draft: T.J.P., K.V., and A.W. Writing—review and editing: T.J.P., K.V., and AW. **Competing interests:** T.J.P 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. K.V. and A.W. declare that they have no competing interests.

Submitted 1 December 2023
Accepted 1 October 2024
Published 30 October 2024
10.1126/scirobotics.adn2733

Science Robotics

Understanding the sense of self through robotics

Tony J. Prescott, Kai Vogeley, and Agnieszka Wykowska

Sci. Robot. **9** (95), eadn2733. DOI: 10.1126/scirobotics.adn2733

Science Robotics (ISSN 2470-9476) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Robotics* is a registered trademark of AAAS.