frontiers | Frontiers in Human Neuroscience

TYPE Hypothesis and Theory
PUBLISHED 18 November 2024
DOI 10.3389/fnhum.2024.1441931

OPEN ACCESS

EDITED BY
Klaus Kessler,
University College Dublin, Ireland

REVIEWED BY
Josselin Baumard,
EA7475 Centre de Recherches sur les Fonctionnements et Dysfonctionnements Psychologiques (CRFDP), France
Marco D'Alessandro,
Neuraptic AI, Spain
Ala Yankouskaya,
Bournemouth University, United Kingdom

*CORRESPONDENCE
Mateusz Woźniak
✉ mgwozniak@gmail.com;
✉ mateusz.wozniak@iit.it

RECEIVED 31 May 2024
ACCEPTED 31 October 2024
PUBLISHED 18 November 2024

CITATION
Woźniak M (2024) How to grow a self: development of self-representation in the Bayesian brain.
Front. Hum. Neurosci. 18:1441931.
doi: 10.3389/fnhum.2024.1441931

How to grow a self: development of self-representation in the Bayesian brain

Mateusz Woźniak[1,2,3,4]*

[1]Social Cognition in Human-Robot Interaction Group, Italian Institute of Technology, Genoa, Italy, [2]Social Mind Center, Department of Cognitive Science, Central European University, Vienna, Austria, [3]Cognition and Philosophy Lab, Department of Philosophy, Monash University, Melbourne, VIC, Australia, [4]Institute of Psychology, Jagiellonian University, Cracow, Poland

The last two decades saw multiple attempts to explain how the self is represented in the brain within the framework of the Bayesian brain. However, these attempts largely focused on describing a developed, adult self-representation. The current paper argues that looking at the developmental trajectory is crucial for understanding the structure of self-representation. It argues that emergence of self-representations should be understood as an instance of the process of acquisition of new internal models of hidden causes of sensory input. The paper proposes how such models emerge and develop over the course of life by looking at different stages of development of bodily and extra-bodily self-representations. It argues that the self arises gradually in a series of discrete steps: from first-person multisensory representations of one's body to third-person multisensory body representation, and from basic forms of the extended and social selves to progressively more complex forms of abstract self-representation. It discusses how each of them might emerge based on domain-general learning mechanisms, while also taking into account the potential role of innate representations. Finally, it discusses how predictions of the proposed model might be experimentally tested.

KEYWORDS

self, self-representation, Bayesian brain, predictive coding, cognitive development, self-recognition, bodily self, abstract self

1 Introduction

What is the self? William James (1890) revolutionized the discussion about this question and set foundations for scientific study of the self by introducing the distinction between self-as-subject ("I") and the self-as-object ("Me"). The importance of this idea lies in the fact that while the problem of "I" is a metaphysical problem (James, 1890; Woźniak, 2018), the "Me" can be investigated as either a type of experience, or as an underlying structure of mental representations (self-models).[1] Arguably, this insight shifted the problem of the self from being a purely philosophical issue into a scientific research question which can be approached by empirical science through investigation of neural and cognitive underpinnings of our self-representations. Afterwards, research on the self entered the era during which it has been investigated as a psychodynamical mechanism (Freud, 1933), symbolic knowledge

[1] For different recent conceptualisations of self-as-object see: Sui, J. and Gu, X. (2017). Self as Object: Emerging Trends in Self Research. *Trends Neurosci*, *40*(11), 643–653. https://doi.org/10.1016/j.tins.2017.09.002 and Truong, G. and Todd, R. M. (2017). SOAP Opera: Self as Object and Agent in Prioritizing Attention. *J Cogn Neurosci*, *29*(6), 937–952. https://doi.org/10.1162/jocn_a_01083

(Baumeister, 1999; Markus, 1977; Mead, 1934; Tajfel, 1982), and a connectionist network (Kihlstrom et al., 2003; Smith et al., 1999).

During the last decade a number of proposals attempted to explain the self as a neurocognitive structure within the predictive (Bayesian) brain. These theories typically (but not always) refer to more specific frameworks of predictive coding and the free energy principle (Friston, 2005, 2010) and target several important aspects of the self, such as body representation (Apps and Tsakiris, 2014; De Vignemont, 2010; Hohwy, 2007, 2013; Limanowski, 2022; Limanowski and Blankenburg, 2013; Salomon, 2017; Seth, 2013), the role of interoception (Allen and Tsakiris, 2018; Fotopoulou and Tsakiris, 2017; Seth, 2013; Seth and Tsakiris, 2018), abstract and social aspects of self-knowledge (Allen and Tsakiris, 2018; Bolis and Schilbach, 2020; Fotopoulou, 2012; Fotopoulou and Tsakiris, 2017; Friston and Frith, 2015; Hohwy and Michael, 2017; Moutoussis et al., 2014), as well as the self as present in conscious experience (Ciaunica, Constant, et al., 2021; Ciaunica, Safron, et al., 2021; Letheby and Gerrans, 2017; Seth et al., 2011; Woźniak, 2018).

Most Bayesian notions of the self focus on the adult self. They describe the structure and dynamics of something that has been already formed throughout development. As such, they tend to omit an important aspect of the problem – the question of how the self emerges and changes during life. However, in the recent years several papers have introduced the developmental perspective into this field. Frederique de Vignemont (2010) tackled developmental issues when discussing emergence of body representations and Anna Ciaunica and colleagues recently stressed the importance of looking at the development when explaining the origins of self-consciousness (Ciaunica et al., 2021a; Ciaunica and Crucianelli, 2019; Ciaunica et al., 2021b). Katarina Fotopoulou and Manos Tsakiris used the developmental perspective to illuminate specific aspects of development of the self such as self-recognition (Drysdale and Tsakiris, 2021) and the role of "mentalization" in constituting one's self (Fotopoulou and Tsakiris, 2017). Finally, Riva (2018) has recently proposed a comprehensive developmental model of bodily self-representation in which he outlined a series of stages during which an adult body-representation comes into being.

This paper aims to introduce a different, partially complementary and partially competing, perspective into this field and propose a mechanistic model of development of self-representation within the framework of the Bayesian brain. The first goal of this paper is to discuss a developmental mechanism of acquisition of new internal models. The second goal is to use it to propose a potential trajectory of development of an adult self-representation. The resulting proposal bears many similarities to Riva's (2018) model and the final section will discuss differences and similarities between these two (as well as other) theories.

The current proposal focuses on self-representation (how our mind/brain represents ourselves) and not on self-consciousness (how we consciously experience ourselves). It takes a representationalist approach (for discussion of what is a representation in cognitive science and predictive coding theories see: Baker et al., 2022; Egan, 2012; Ramsey, 2016; Williams, 2018) and assumes that self-representations underpin conscious self-experience, but are not synonymous with it, as many self-representations are unconscious. On the other hand, this approach also assumes that any self-related conscious experience must be underpinned by some form of self-representation, meaning that research on self-consciousness remains directly relevant to the discussion of self-representation. Finally, this paper limits itself to presentation of the theory at the computational level (Marr, 1982), although it also refers to recent developments which can illuminate it on algorithmic and implementation levels.

I will describe the model mostly through the lens of the predictive coding framework (Friston, 2005; Rao and Ballard, 1999; Spratling, 2016, 2017, for discussion of neural implementation see: Bastos et al., 2012; Keller and Mrsic-Flogel, 2018; Shipp, 2016), although it is equally compatible with other Bayesian and predictive models of cognition (Fiser et al., 2010; Griffiths et al., 2010; Knill and Pouget, 2004; Orbán et al., 2016; Tenenbaum et al., 2011, also: Heeger, 2017), as well as with connectionist models of cognition (DiCarlo et al., 2012; Kriegeskorte, 2015; Kumaran et al., 2016; McClelland et al., 2010; McClelland and Rogers, 2003; Yamins and DiCarlo, 2016).

The paper is divided into two parts. The first part will clarify the theoretical background by (1) specifying what I here will understand as self-representation, (2) describing basic postulates of the Bayesian model of the mind, and (3) explaining how to understand cognitive development within the Bayesian framework. The second part of the paper will then propose a model of cognitive development of self-representation.

2 Self-representation, Bayesian brain, and acquisition of new models

2.1 The self as a representational structure

The goal of this paper is to propose a theory of development of the self understood as self-representation Bayesian approaches often assume a representational view of the mind (Hohwy, 2013, but see: Clark, 2016; Williams, 2018). In line with this assumption, when I speak about the self, I refer to a representational structure, which is encoded in the brain. Aspects of this structure can become conscious, but in principle mental representations can be analyzed independently of whether they are conscious or not (Chalmers, 2004; Crane, 2003).

This understanding of the self has been implicitly assumed in much of traditional research in cognitive science and psychology, especially within the connectionist framework, in which self-representation is understood as a structure in memory (e.g., Kihlstrom et al., 2003; Smith et al., 1999, but also: Conway, 2005; Hart et al., 1997; Haslam et al., 2011; Martinelli et al., 2013). The connectionist approach can provide a useful way of conceptualizing self-representation as a network composed of: (a) nodes representing memory content, which are linked to (b) a node representing the internal model of the self (Figure 1). Under such conceptualization, what makes my mental representation of my face a part of self-representation is the fact that it is linked with my self-model, while my representation of another face (my friend's) is not. A connectionist network provides a convenient way to visualize all aspects of the self-representation on a two-dimensional plane. However, as will be argued later, the structure of the self is more complex.

2.2 The self as an internal model

In Bayesian and predictive models the brain is seen as an inference machine which continuously attempts to explain the world by combining sensory input (data in Bayesian terms, or prediction error in predictive coding framework) with prior knowledge (priors or predictions, respectively) in order to yield a posterior representation (or more properly: posterior distribution) which is responsible for the

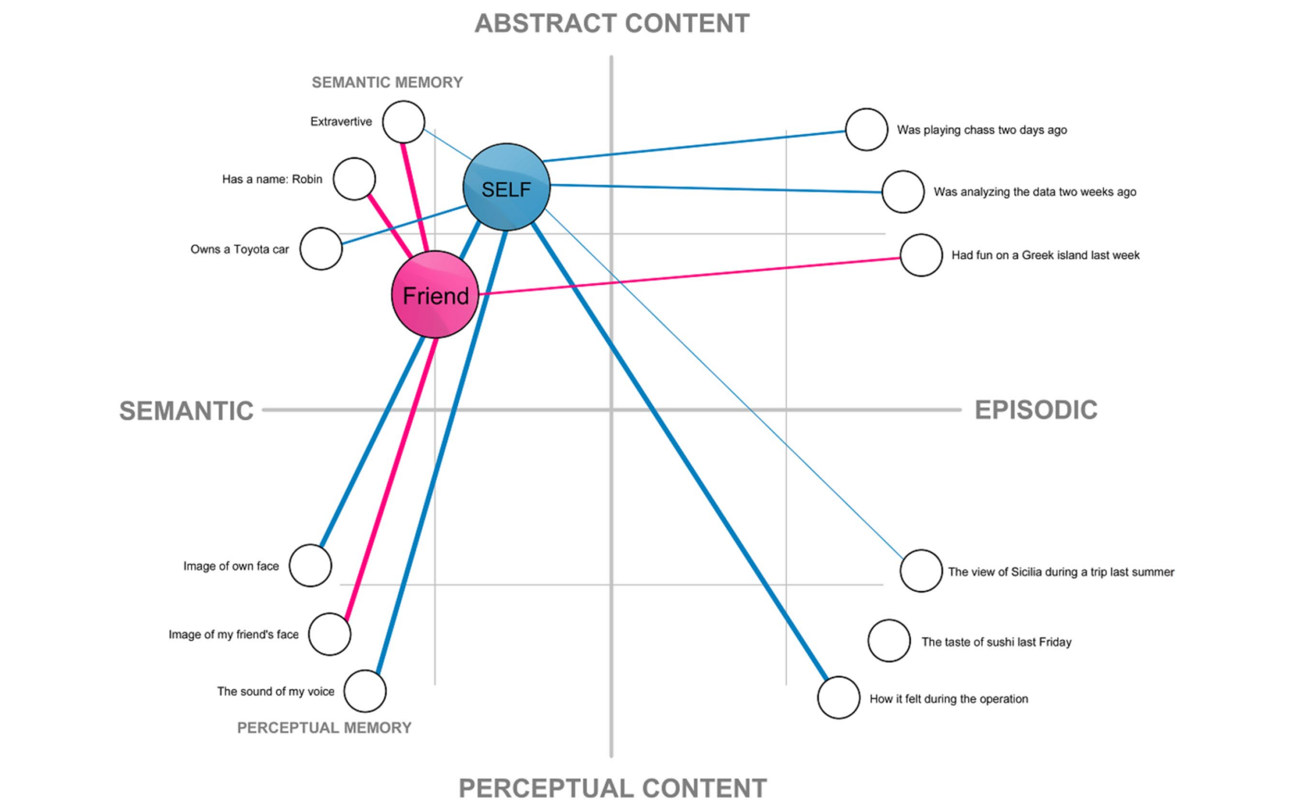
ABSTRACT CONTENT
SEMANTIC MEMORY
Extravertive
Has a name: Robin
Owns a Toyota car
SELF
Friend
Was playing chass two days ago
Was analyzing the data two weeks ago
Had fun on a Greek island last week
SEMANTIC
EPISODIC
Image of own face
Image of my friend's face
The sound of my voice
PERCEPTUAL MEMORY
The view of Sicilia during a trip last summer
The taste of sushi last Friday
How it felt during the operation
PERCEPTUAL CONTENT

FIGURE 1
An example of a localist connectionist model of self-representation based on the associative network model of memory (cf. Kihlstrom et al., 2003; Smith et al., 1999). Mental representations are presented as nodes on a two dimensional plane spanning perceptual-to-abstract content and semantic-to-episodic content. The internal model of the self and a friend are visualized as elements of the semantic memory. These nodes are connected with other representations. Thicker lines denote strong connections between the nodes, while thinner lines represent weaker ones.

eventual percept. It can be illustrated with an example. Let us say that I invited two of my friends, Anna and Julia, for dinner at 7 pm. Exactly at 7 pm I hear a door bell. I know that Anna always comes on time, while Julia is well-known for being late. Given this prior knowledge, I am certain that it is Anna waiting at the door. I open the door and it takes me a while to realize that I'm standing in front of Julia instead. Under Bayesian interpretation, I had a very strong prior belief that I will encounter Anna. However, the sensory input strongly suggested otherwise because I saw a person looking like Julia at the door. Sensory data eventually overcame my prior belief and led to the formation of a posterior belief that it was Julia that came first, and not Anna.

Conducting this type of (Bayesian) inference requires having a hypothesis space. In its simplest form it can be just a collection of individual hypotheses (a categorical distribution). In the example above there were only two hypotheses: that a person at the door is Anna or Julia, but one can entertain more hypotheses, e.g., that it is a postman or a neighbor[2]. Hypothesis spaces can also have a continuous form, e.g., when I estimate the height of a person (what can be parametrized as a normal distribution) or somebody's number of Facebook friends (typically following a power-law distribution). Regardless of the form of the hypothesis space, in order to calculate the posterior distribution using Bayesian inference, I need to have models (representations) of possible causes. For example, if I open the door and see a wallaby there, I will not be able to recognize it as what it is, until I possess an internal model of a wallaby. If I do not have such model then I may try to accommodate what I see into the closest of my existing models - in this case, for example, my internal model of a kangaroo.

[2] In principle a categorical hypothesis space might contain almost an infinitely large number of categories. This might raise a concern that such mechanism might be computationally inefficient or even intractable. One solution to this problem has been provided by sampling-based approach to Bayesian cognition. Under this view we do not perform Bayesian inference over the full hypothesis space at once, but rather sample from it across time. It means that we might still entertain a very unlikely hypothesis if we spend enough time exploring the hypothesis space. This might happen if our sensory data is highly different than our prior distribution (what I see is very different from what I expected to observe) and I keep trying to find a satisfying explanation. For example, if I see a cat floating in the air in front of me I might try to explain it by going through otherwise unlikely explanations: am I dreaming? Am I in the Matrix? Was my food spiked with drugs? For sampling based approaches see: Fiser, J., Berkes, P., Orbán, G., and Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences*, *14*(3), 119–130. Orbán, G., Berkes, P., Fiser, J., and Lengyel, M. (2016). Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron*, *92*(2), 530–543.

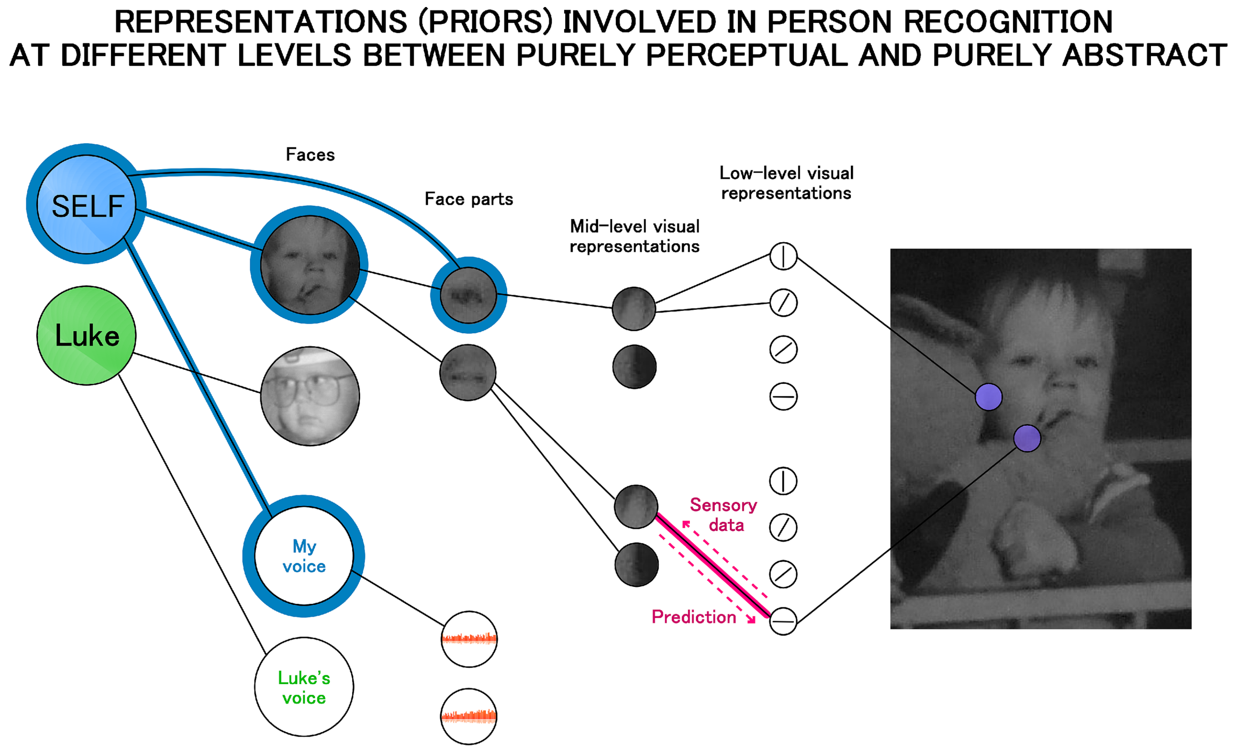
REPRESENTATIONS (PRIORS) INVOLVED IN PERSON RECOGNITION
AT DIFFERENT LEVELS BETWEEN PURELY PERCEPTUAL AND PURELY ABSTRACT
Faces
Face parts
Low–level visual representations
Mid–level visual representations
SELF
Luke
My voice
Luke's voice
Sensory data
Prediction

FIGURE 2
Self-representation as a hierarchical structure of internal models with the self-model as a high level abstract prior (on the left). Each link in the hierarchical structure reflects the interplay of top-down and bottom-up signals as illustrated by the link in pink. In Bayesian terms the top-down signal reflects prediction or strength of the prior, and the bottom-up direction reflects prediction error or the incoming data. The self-model (the blue node) is connected to the representation of a person's voice and face, but also to the representation of one's eye illustrating that representations are free to be connected across levels of the hierarchy.

Under predictive coding and other Bayesian models the brain is understood as a network of internal (generative) models representing hidden causes of sensory input. This network is organized in a hierarchical manner. The most abstract causes are at the top (high-level, including social, semantic categories which are represented in the anterior temporal lobes, orbitofrontal cortex, and ventral medial prefrontal cortex: Bowman and Zeithamova, 2018; Chen et al., 2017; Clarke and Tyler, 2015; Ralph et al., 2017) and the most basic perceptual ones at the bottom (e.g., models representing line orientations on a patch of the retinal input are represented in primary visual cortex: Hubel and Wiesel, 1962, 1968). According to some predictive coding theories the whole brain is organized in this way (Friston, 2010), while others attribute this architecture only to the cortex (e.g., Friston, 2005; Spratling, 2010). Nevertheless, both views imply that self-representation must be underpinned by this architecture as well, and the theory proposed here shares this assumption. Figure 2 illustrates it graphically. It provides an elaboration of the connectionist conceptualization of the self, but also includes visualization of the hierarchical structure of internal models (representations).

2.3 Learning new internal models

Across lifespan people not only use representations with which they were born (for a discussion see: Bottari et al., 2015; Carey, 2009; Heyes, 2018; Reid et al., 2017), but also acquire new ones. In the Bayesian framework this raises the problem of how we come to acquire new internal models (representations). If I see a person in front of me I will assume that I am facing a living human being. However, if I possess an internal model of a TV then I may be capable of forming an alternative hypothesis – that the human figure in front of me is not a real person, but just a video recording played on a TV screen. Acquiring new models allows us to expand the hypothesis space and, as a consequence, allows more hidden causes to be taken into account when trying to make sense of a given situation.

Susan Carey (2009) has proposed an influential theory of development of understanding of concepts. While she focused on concepts, her proposal can be applied to development of new internal models in general. In her view conceptual development is a discontinuous process consisting of episodes of qualitative change, during which the existing concepts (representations) are recombined into new, more powerful ones.[3] She illustrates it with an example of acquiring the concept of a natural number by human infants. Until 2 years old, most children do not seem to understand the concept of a number at all. Then, usually between 24 and 30 months of age, they begin to understand the concept of "one." Approximately 6 to 9 months

[3] The origins of this idea can be traced back to the theory of assimilation and accommodation introduced by Jean Piaget Piaget, J. (2013). *The construction of reality in the child*. Routledge.

later they begin to grasp the concept of two, but still fail with larger numbers. Later on, they become "three-knowers," and sometimes also "four-knowers," but around that time most of them undergo a qualitative change in their conceptual representation of numerosity, which leads them to understanding of the concept of a natural number (what happens typically when they are between 3 and 6 years old). Carey explains this transition by proposing that children recombine their primitive representations (structures which are part of the innate core knowledge, in this case the "parallel individuation system" and the "analog magnitude system") through a process called "Quinian bootstrapping" (Figure 3). It is not necessary to commit to the existence of an innate stock of representations, nor to the claim that conceptual change can happen only through Quinian bootstrapping, to grasp the main idea behind her proposal: we acquire new concepts by recombining pre-existing representations, what leads to the emergence of new mental representations which are incommensurable with the old ones. In case of the natural numbers: we begin with understanding of the concept of "one," then "two" and "three," and then at some point we discover the underlying rule by mapping them onto our representation of magnitude. The discovery of that rule leads to mental reorganization which translates to discovery of the concept of a natural number. At later stages children and adults can recombine this concept of a natural number with other concepts to acquire even more abstract concepts like integer or rational number, and even to revisit the concept of a natural number from a different perspective, e.g., when deliberating whether zero is a natural number or not.

If we apply Carey's idea to the Bayesian brain framework, then development of new internal models can be understood as the process of restructuring of a hypothesis space leading to the emergence of new internal models (cf. Perfors et al., 2011; Tenenbaum et al., 2011; Ullman and Tenenbaum, 2020) This idea has been recently developed in computational models attempting to model the process of discovery of new categories (Love et al., 2004; Navarro and Kemp, 2017; Pothos and Chater, 2002; Smith et al., 2019). One prominent example of such process has been illustrated in models of learning of the structural form of data (Kemp and Tenenbaum, 2008; Lake et al., 2018), but for our purposes a more relevant example has been provided by Hohwy and Michael (2017), who wrote about a similar process in the context of self-recognition. Hohwy and Michael described a situation of a human organism, which detects brief periods of darkness every few seconds. One possible cause of this situation is that the light goes off and on. But there is an alternative explanation once one realizes that brief periods of darkness are perfectly correlated with motor commands sent to one's eye lids. As such, one may come up with a second explanation, that it's not the world, but oneself that is switching the light on and off by opening and closing the eyes! However, in order to take this hypothesis into consideration one needs to possess an internal model of oneself as a hidden cause of sensory input, i.e., a model of the self. In other words, in order for me to be able to realize that a hidden cause of my sensory input is "me," I need to possess at least a rudimentary model that such a hidden cause as "me" exists. The remaining part of the paper will introduce a proposal of how such a model emerges and how it develops across lifespan (Figure 3).

3 Development of the self

Previous sections introduced a picture of the brain as a Bayesian inference machine and described the place of the self in its architecture. It can be summarized as follows:

1. The brain can be understood as a Bayesian hierarchical inference machine, which attempts to model the environment by inferring hidden causes of sensory input. The self-model is just one model of such hidden causes.
2. Self-representation can be understood as a network composed of (a) a central node representing the self-model (which is a representation of a hidden cause, i.e., oneself), linked with (b) multiple other representations (other hierarchical representations of hidden causes, e.g., own arm, own hand, own finger, one's nationality, one's autobiographic memories, etc.).
3. Cognitive development is a process of acquiring new representations, i.e., new internal models, through a process of recombination of the existing representations. Here, I do not specify the mechanism which allows the brain to do it, but I will suggest some of them when discussing individual developmental steps.

Based on these assumptions the rest of the paper will outline a proposal of the developmental trajectory of the self, which is understood as a representational structure composed of multiple internal models. I will argue that the adult human self-representation emerges as a consequence of a series of discrete developmental steps, beginning with the emergence of a primordial form of the self-model and ending with a mature form of an abstract self-model. Figure 4 provides a brief summary of the main postulated developmental steps.

3.1 The primordial self: sensorimotor and interoceptive body representations

It is typically assumed that a person is born after leaving the womb. However, birth follows a 9-month long period of intensive prenatal development during which the neural system is being formed and organized (Anderson and Thomason, 2013; Kostović and Judaš, 2010; Moore and Linthicum Jr, 2007). Moreover, even before birth foetuses show a wide range of spontaneous behaviors, react to stimulation in a wide range of sensory modalities, and even exhibit behavioral displays of learning (Anderson and Thomason, 2013; Hepper, 2015). Given all of these, a full account of cognitive development, including development of the self, must also take into account processes which take place during the foetal period of life (Ciaunica et al., 2021a; Ciaunica et al., 2021b).

Can a primordial self-model emerge during prenatal development? Previous theories postulating the importance of sensorimotor integration for the emergence of the self have suggested how such a process might take place (although within a non-representationalist framework: Christoff et al., 2011; Legrand and Ruby, 2009). Let us assume an idealized scenario in which a foetus has already developed some basic sensory (tactile, proprioceptive, nociceptive, visceral) systems which allow their brain to detect sensory stimulation (without claiming that it happens consciously). If we idealize and assume that at that point their brain does not yet have any models of external causes of this stimulation, we may compare this situation to experiencing a multisensory perceptual noise – something similar to hearing a uniform white noise signal, but across many sensory modalities. It is a situation in which one cannot make sense of anything in the environment, because one does not have any models of what can be out there. The

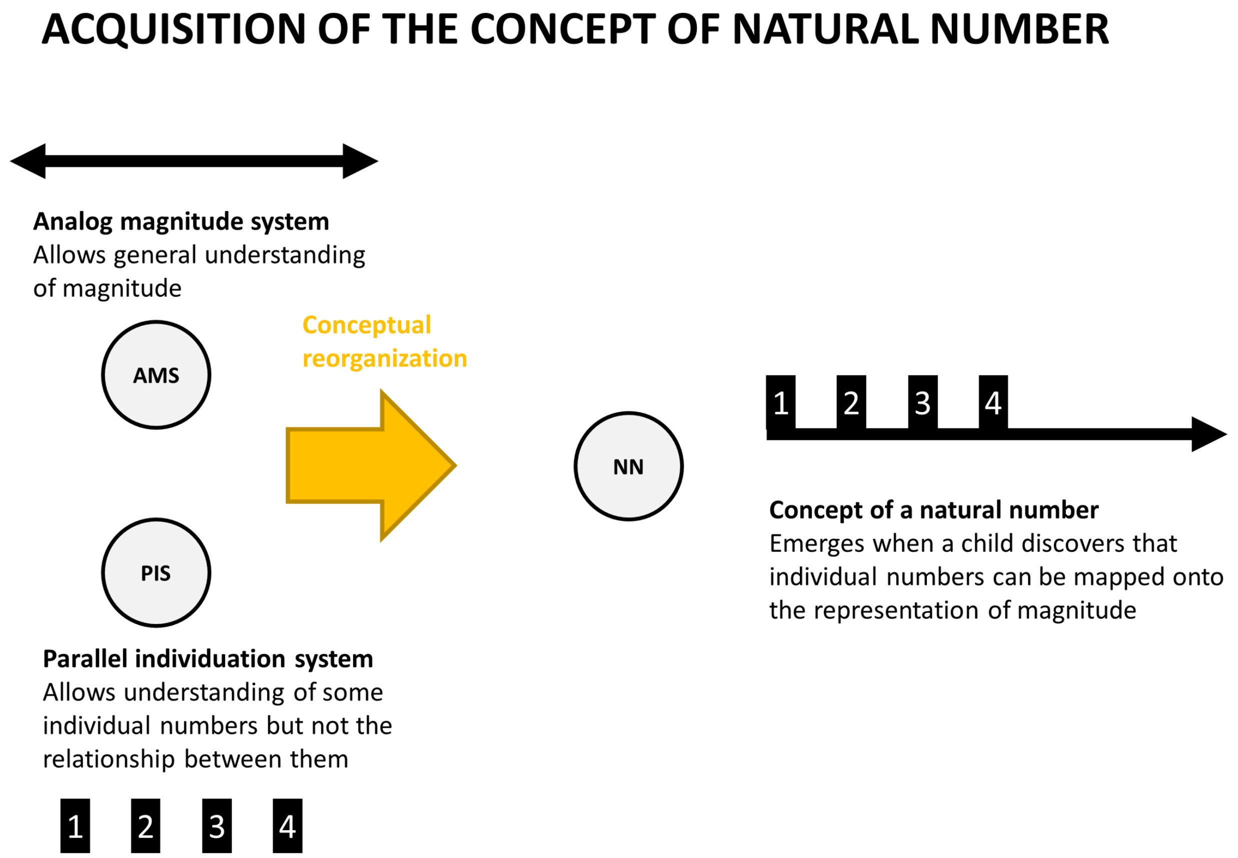
ACQUISITION OF THE CONCEPT OF NATURAL NUMBER
Analog magnitude system
Allows general understanding of magnitude
AMS
Conceptual reorganization
NN
1
2
3
4
Concept of a natural number
Emerges when a child discovers that individual numbers can be mapped onto the representation of magnitude
PIS
Parallel individuation system
Allows understanding of some individual numbers but not the relationship between them
1
2
3
4

FIGURE 3
A graphical illustration of the process of acquisition of the concept of a natural number (NN). According to Carey (2009) children are born with an innate concept of magnitude (through "Analog Magnitude System," or AMS) which allows them to differentiate between, e.g., long and short objects. Children are also born with a "Parallel Individuation System" through which during the third year of life children begin to understand the concept of individual numbers: one, two, three, sometimes four. The crucial point of Carey's proposal is that at certain point children discover that individual numbers can be mapped onto the underlying representation of magnitude (Carey proposes that it happens through the mechanism of Quinian bootstrapping) and hence acquire the concept of a natural number.

only thing that one has is an implicit model to account for everything that happens in the environment: the model of sensory noise (Figure 5A). Second, let us also assume that the foetus has a rudimentary motor system, which allows them to initiate random movements (indeed foetuses exhibit a rich repertoire of motor behaviors: Arduini et al., 2013; DiPietro et al., 1998; James et al., 2013; Robertson, 1990). Each such movement will be accompanied by some pattern of sensory consequences. For example, performing a random hand movement will lead to a specific tactile and proprioceptive pattern of sensations, and performing a leg movement will lead to a different pattern. In both cases these patterns will be structured in a different way than any pattern of stimulation elicited by external sources, such as movements of their mother's body. However, the pattern of sensory stimulation accompanying one's movements (that will come as a consequence of initiating a motor command) will covary with one's movement with much greater regularity than the pattern of stimulation caused by the mother. This constitutes a reinterpretation of the notion of self-specifying processes discussed by Christoff et al. (2011). Similar models have been also proposed in developmental robotics (Hoffmann, 2022; Schillaci et al., 2016).

At this point the task of the brain is to pick up on this regularity and perform something akin to cluster analysis. It needs to utilize the statistical regularity between motor commands (or more specifically: patterns of neural activation in the motor system, even if they are generated randomly) and their sensory consequences (patterns of neural activation in the sensory systems) and discover that it can probabilistically classify the sensory signal into two categories: something that can, and something that cannot be predicted by motor activity. This might result in the emergence of two categories at the top level of the sensory hierarchy, and any sensation can thereafter be classified as one of these two: something caused by me or something caused by something else. Although the foetus will likely not be aware of the meaning of this classification, it will mark the emergence of the fundamental distinction between the self and the environment, and representations of the corresponding categories can be regarded as primordial models of the self and environment.

Within the Bayesian framework this situation can be rephrased as a simple instance of formation of models of two hidden causes,[4] based on statistical regularities detected in the sensory input. As such, it can be regarded as one of the simplest instances of new model acquisition

[4] It is likely that in practice only the explicit model of the self emerges, and the non-self model functions as an implicit model representing all the unexplained residual "noise."

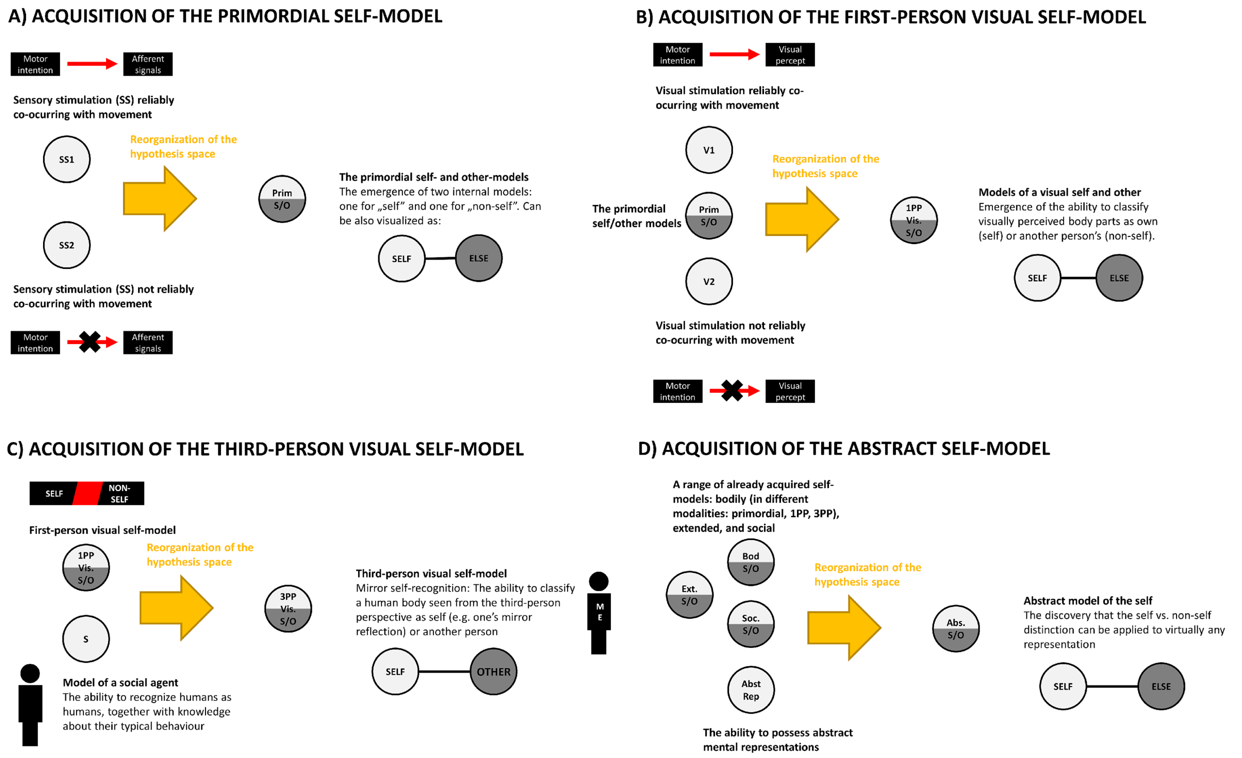
A) ACQUISITION OF THE PRIMORDIAL SELF-MODEL
Motor intention
Afferent signals
Sensory stimulation (SS) reliably co-ocurring with movement
SS1
Reorganization of the hypothesis space
Prim S/O
The primordial self- and other-models
The emergence of two internal models: one for „self" and one for „non-self". Can be also visualized as:
SELF
ELSE
SS2
Sensory stimulation (SS) not reliably co-ocurring with movement
Motor intention
Afferent signals
B) ACQUISITION OF THE FIRST-PERSON VISUAL SELF-MODEL
Motor intention
Visual percept
Visual stimulation reliably co-ocurring with movement
V1
Reorganization of the hypothesis space
The primordial self/other models
Prim S/O
1PP Vis. S/O
Models of a visual self and other
Emergence of the ability to classify visually perceived body parts as own (self) or another person's (non-self).
SELF
ELSE
V2
Visual stimulation not reliably co-ocurring with movement
Motor intention
Visual percept
C) ACQUISITION OF THE THIRD-PERSON VISUAL SELF-MODEL
SELF
NON-SELF
First-person visual self-model
1PP Vis. S/O
Reorganization of the hypothesis space
3PP Vis. S/O
Third-person visual self-model
Mirror self-recognition: The ability to classify a human body seen from the third-person perspective as self (e.g. one's mirror reflection) or another person
ME
S
SELF
OTHER
Model of a social agent
The ability to recognize humans as humans, together with knowledge about their typical behaviour
D) ACQUISITION OF THE ABSTRACT SELF-MODEL
A range of already acquired self-models: bodily (in different modalities: primordial, 1PP, 3PP), extended, and social
Bod S/O
Ext. S/O
Soc. S/O
Reorganization of the hypothesis space
Abs. S/O
Abstract model of the self
The discovery that the self vs. non-self distinction can be applied to virtually any representation
Abst Rep
SELF
ELSE
The ability to possess abstract mental representations

FIGURE 4
Graphical illustration of acquisition of four types of self-models. In each case the existing representations/models (shown on the left in each section) are recombined (indicated by the yellow arrows) into novel self-models (shown on the right). **(A)** The primordial self-model emerges once the brain discovers that it can classify sensory signals into the ones which are reliably predicted by one's motor behaviour and those which are not. **(B)** The same classification rule can be applied to visual perception leading to acquisition of first-person representation of one's body. **(C)** A toddler that possesses the concept of another person and understands that some visual percepts represent their body can discover that in special cases (e.g., encounter with a mirror) they can come together, and that it is possible to see oneself from the third-person perspective. **(D)** An abstract self-model emerges once a child discovers that self/non-self distinction can be applied to abstract representations.

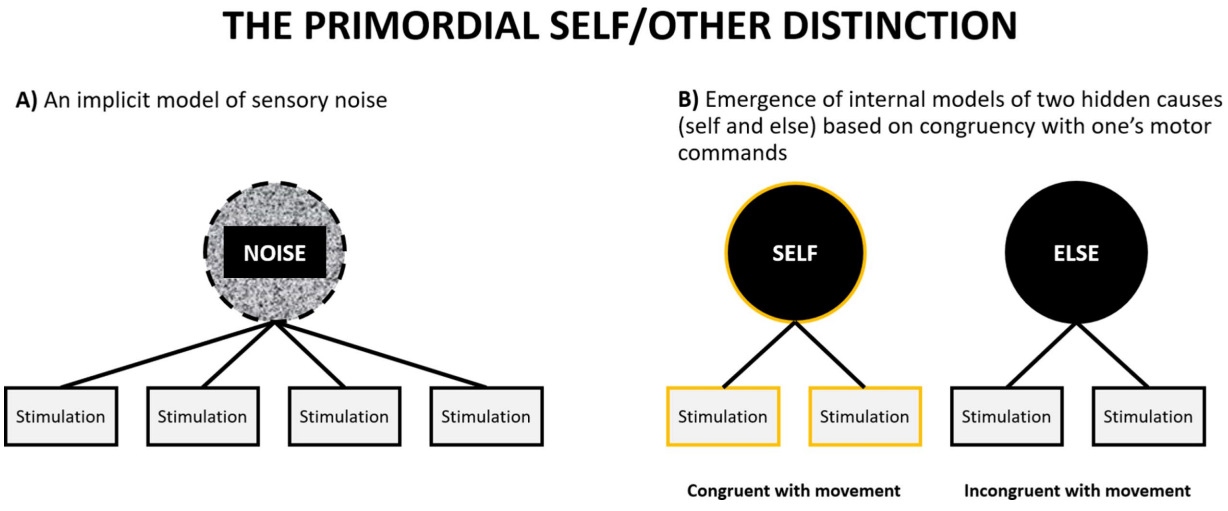
THE PRIMORDIAL SELF/OTHER DISTINCTION
A) An implicit model of sensory noise
NOISE
Stimulation
Stimulation
Stimulation
Stimulation
B) Emergence of internal models of two hidden causes (self and else) based on congruency with one's motor commands
SELF
ELSE
Stimulation
Stimulation
Stimulation
Stimulation
Congruent with movement
Incongruent with movement

FIGURE 5
A schematic illustration of emergence of the primordial self-model. **(A)** At the outset an organism has only an implicit model of sensory noise. **(B)** The discovery that some sensory stimulation reliably co-occurs with one's motor commands leads to the emergence of the primordial models of self and else (see also Figure 4A).

and is in agreement with Bayesian computational models of the brain (Lake et al., 2018; Perfors et al., 2011; Tenenbaum et al., 2011; Ullman and Tenenbaum, 2020). However, is the prenatal brain capable of implementing such an operation? It appears that yes. Twins *in utero* show developmental differences between touching behavior toward themselves, the other twin, and the uterine wall (Castiello et al., 2010). These results suggest that foetuses can not only discriminate between self and the uterine environment, but also between self and another foetus. There is also evidence that foetuses can learn to discriminate between the voice of mother and other people (DeCasper and Fifer, 1980; DeCasper et al., 1994; Kisilevsky et al., 2003), what presents a much more difficult problem requiring much finer discrimination than a simple ability to distinguish between self-caused and externally caused sensations. Second, computational models of human brain can easily learn much more sophisticated multidimensional discriminations (e.g., Lake et al., 2018; Smith et al., 2019). For example, based on earlier computational models it has been argued that emergence of receptive fields in the visual cortex can be understood as an instance of unsupervised learning over visual input, which leads to performance of something akin to principal components analysis (Olshausen and Field, 1996; Simoncelli and Olshausen, 2001). In this context, forming two models reflecting self and other seems to be a trivial task, even for a brain at a very early stage of development.

It is difficult to overestimate the usefulness for survival of being able to perform this distinction. Therefore, it is highly plausible that most living organisms which possess nervous system are evolutionarily pre-equipped with it. Regardless of whether it is acquired through learning or as a part of innate core cognition, it appears that a newborn child enters the world with a basic ability to differentiate between self and everything else, an ability which, under the Bayesian account, is underpinned by existence of specific models of self and else.

This view of the primordial self builds on and is in agreement with theories proposed previously by other authors which emphasize that the self emerges on the basis of sensorimotor signals (Blanke, 2012; Blanke and Metzinger, 2009; de Klerk et al., 2021; Gallagher, 2000; Gallagher, 2006; Salomon, 2017), and especially the ones which treat the sensorimotor loop as a self-specifying process (Christoff et al., 2011; Legrand and Ruby, 2009). Just like these previous proposals it postulates that sensorimotor congruency is fundamental for the emergence of the most basic form of self-representation. It is also in partial agreement with theories that draw attention to the role of interoception for the self (Allen and Tsakiris, 2018; Babo-Rebelo and Tallon-Baudry, 2018; Damasio, 1999; Park and Tallon-Baudry, 2014; Seth, 2013; Seth et al., 2011; Seth and Tsakiris, 2018), although in the proposed view the role of interoception is reduced. Interoception typically operates on a longer temporal scale, and does not provide as clear pattern of statistical regularities differentiating between self- and externally-generated actions as motor behavior, making it a worse candidate for bootstrapping the primordial self-other distinction. However, it is likely that once the brain discovers this fundamental distinction through sensorimotor congruence, interoception becomes classified as an internal process, and as such becomes a constituting part of the representational self.

3.2 The auditory and first-person visual body representation

At the moment of birth, one's stream of sensory stimulation changes. Once out of the womb, one begins to receive much more detailed visual input, while at the same time one's tactile input drastically changes its nature (one is no longer submerged in uterine fluid). In regard to vision a foetus can only detect big differences in luminance (Eswaran et al., 2004; Peleg and Goldman, 1980), while a newborn becomes exposed to a wide variety of complex visual scenes. However, the scope of possible experiences increases in all sensory modalities, not only in vision.

This explosive increase of richness of experienced sensations means more data that the brain needs to make sense of. And it includes the possibility to classify elements of this sensory stream into self-related versus non-self-related. Importantly, a newborn needs to learn to distinguish between self- and other-generated sounds (e.g., crying), and between seen objects which are parts of the environment versus the ones which are parts of one's body. I postulate that in both cases of vision and audition the underlying mechanism of learning to make the distinction between self and non-self will be similar to the one described in the previous section – the brain will need to pick up on the regularities between one's motor activity and its corresponding auditory and visual consequences. The ones which reliably accompany one's motor activity will then become represented as forming parts of one's self. It may be the sound of one's crying, or the image of one's body parts moving as seen from the first-person perspective. Once learned, they become incorporated into the structure of internal representations of hidden causes which are classified as reflecting the self.

How quickly does a newborn learn that they can have agency over seen objects? Several studies suggest that this ability is developed during the first year (Kenward, 2010; Miyazaki et al., 2014; Zaadnoordijk et al., 2020), perhaps as early as in the second month of life (Rochat and Striano, 1999, 2000; Watanabe and Taga, 2006, 2009, 2011; Watson and Ramey, 1972). However, all of these studies investigated signs of agency over objects external to one's body and utilized procedures in which interaction with these objects was very short (typically several minutes). It is possible that a newborn acquires agency over one's seen body parts, especially hands, much earlier, as a consequence of the fact that they are much more reliably associated with one's motor commands than any external object, and that newborns have much more time to learn it (virtually all of the waking life after birth). This possibility is further supported by findings showing that even newborns are sensitive to visuo-tactile congruency (even for images of faces: Filippetti et al., 2013; Filippetti et al., 2015), which is a necessary prerequisite for visual self-recognition.

How do these new visual and auditory self-representations relate to the self-model which has been already developed in a foetus? There are several theoretical possibilities, but the two main ones are: either they get incorporated into the pre-existing self-model or new specialized self-models are being formed (Figure 6). In the first case, the pre-existing self-model becomes expanded to include new auditory and visual aspects. It means that no new model is formed – the old one is modified to accommodate additional modalities. Alternatively, new individual self-models might emerge for each modality or aspect of one's self-representation.

What is important is that these two options lead to different empirical predictions. If each self-model is localized in a different part of the brain then people with lesions in different brain areas should selectively lose individual self-models. For example, it should be possible to find clinical conditions in which one selectively loses an ability to visually recognize one's body from the first person perspective, while retaining the ability to use tactile and prioprioceptive information for self-identification. Although

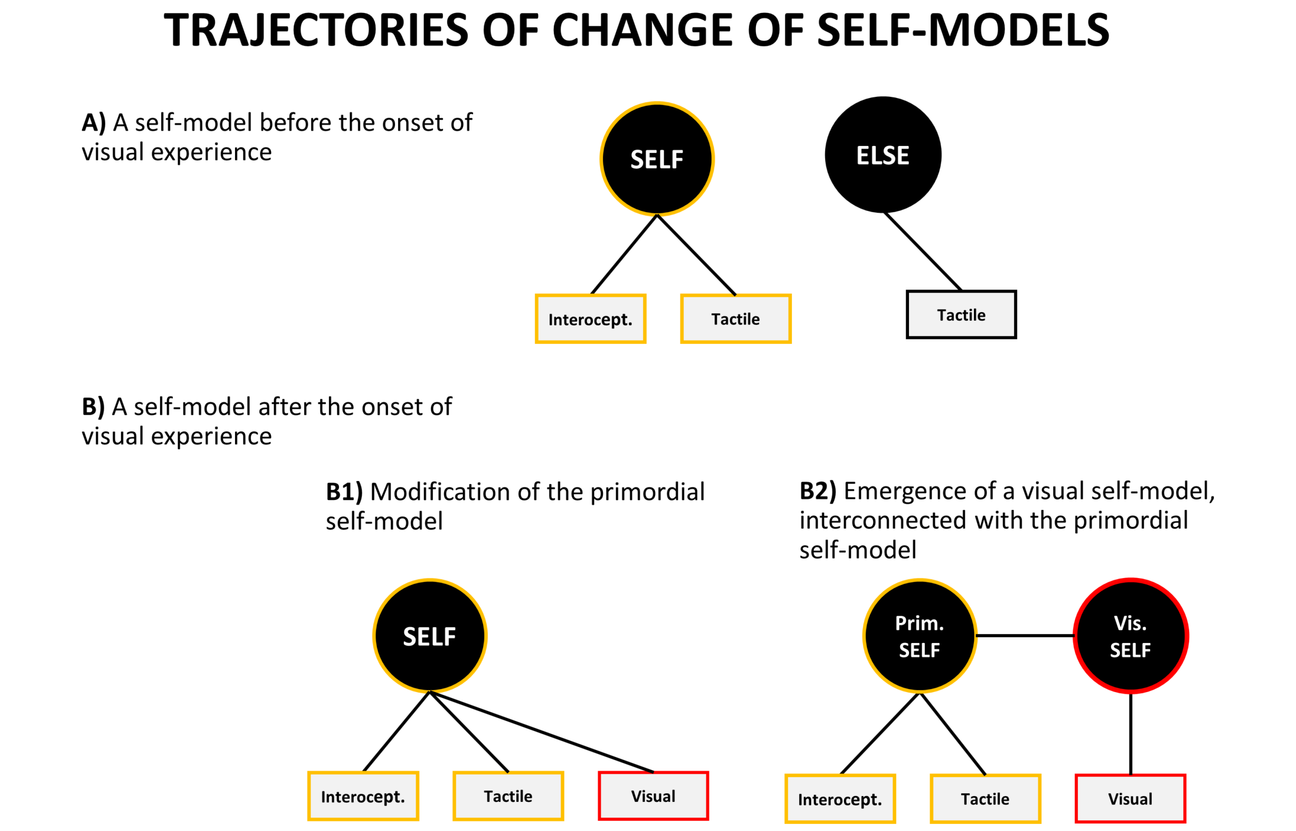
TRAJECTORIES OF CHANGE OF SELF-MODELS
A) A self-model before the onset of visual experience
SELF
ELSE
Interocept.
Tactile
Tactile
B) A self-model after the onset of visual experience
B1) Modification of the primordial self-model
SELF
Interocept.
Tactile
Visual
B2) Emergence of a visual self-model, interconnected with the primordial self-model
Prim. SELF
Vis. SELF
Interocept.
Tactile
Visual

FIGURE 6
Theoretically possible trajectories of change of self-concept. However, as described in the text, empirical evidence favors a variant of model B2.

extremely rare, there are clinical cases providing preliminary indication that this situation might happen. Tobita et al. (1995) reported a case of a 52-year old man suffering from progressive cortico-basal degeneration (which began in the left parietal lobe, which is sometimes implicated as a crucial area for one's body representation) who displayed the following symptoms:

> He [the patient] could not point at any part of his own body in response to verbal or visual commands. On the other hand, he could point at every part of the examiner's body or of the illustrated body image. Deep sensations and linguistic functions were not involved. This cognitive impairment was regarded as autotopagnosia. In contrast with inability to recognize any part of the own body in response to the commands, he could name every part of his body as soon as the examiner touched there. Moreover, his symptoms of autotopagnosia were ameliorated by looking at himself in a mirror; he could point at any part of his own body. Tobita et al. (1995, p. 296).

This case illustrates a condition in which the patient appears to have selectively impaired first-person visual body-representation, while had a preserved third-person visual self-representation (pointing was preserved when looking at one's mirror reflection), and tactile representation (he could name his body parts after being touched).

A similar case, although limited to the left arm, was reported by Verret and Lapresle (1978). They described a patient, claiming that she did not have her left arm. At the same time, she was able to recognize it when touching it with her right hand or when seeing it in a mirror (but only when her left hand was not visible).

It is important to note that in both of the described cases the impairment manifested itself only in regard to the first-person visual representation of one's body, and not for its third-person version – their mirror reflection. This suggests that these two visual representations of our bodies may be underpinned by separate internal self-models. Indeed, there are other reasons to believe that it is the case, including the ones which suggest that the third-person representation of one's body comes as the next step of development of the self. On the other hand, the described case studies by themselves provide only preliminary indication of the full double dissociation that needs to be shown to provide convincing evidence for the proposed model. A full in-depth review of neurological data could provide such evidence – it is however beyond the scope of the current paper.

3.3 The third-person visual body representation

After the first year of life an infant knows at least two things about oneself: (1) that some things that one sees are parts of one's body (forming the first-person visual self-representation), and (2) that some things that one sees are other living creatures, including other people. However, it seems that only until several months later an infant becomes capable to realize that in some specific situations these two things can come together in the form of one's mirror image – something that looks like someone else but in fact is me.

From the point of view of an infant, when seeing one's mirror image for the first time, one encounters a curious creature – something that looks just like any other infant, but is characterized by an uncommon attitude. This attitude manifests itself by the fact that this creature imitates our infant. Moreover, the creature is so skilled in this task that even though an infant tries as hard as possible, they can never surprise the creature by doing something that the latter cannot predict and perform as well. As human adults we know how to make sense of this situation – contrary to the appearance, what one sees in a mirror is not someone else, but oneself as seen from the third-person perspective. However, for an infant finding the solution to this conundrum seems to be much more difficult.

As illustrated by research conducted during the last 50 years, acquisition of the skill to recognize oneself in a mirror (MSR) is a process which takes place slowly and in stages (Amsterdam, 1972; Bertenthal and Fischer, 1978; Filippetti and Tsakiris, 2018; Lewis and Brooks-Gunn, 1979; Lewis and Ramsay, 2004; Schulman and Kaplowitz, 1977). At first (between 6 and 12 months), infants treat their reflection like a potential playmate (Amsterdam, 1972). However, they quickly become frustrated, possibly as a consequence of the fact that the reflection does not enter into meaningful interaction, and eventually begin to ignore the character in the mirror. It is only when they reach the age of between 18 and 24 months when they begin to exhibit clear signs of understanding of the nature of their mirror reflection, such as correctly naming it when asked by adults.

I want to propose that this developmental pattern can be explained as an instance of new model acquisition. At the outset, an infant possesses a visual first-person self-representation (one's body as seen from the first-person perspective) and an internal model of another person (how another person behaves and looks from the third-person perspective). Once an infant encounters a mirror reflection (or any other reflection, e.g., on a water surface), she or he meets an agent which appears to be another person, but at the same time violates many expectations about how another person behaves. This leads to the build up of prediction error, which needs to be explained away. One can do it by simply avoiding mirrors or ignoring them (what corresponds to the active inference strategy: Friston, 2010; Friston et al., 2010), although this strategy will not solve the problem in the long run – one cannot avoid mirrors for the rest of one's life. A better strategy is to either adjust the existing internal model of other people in order to account for such individuals, or to create a new model for this special class of individuals. However, in both of these cases the outcome is suboptimal – the peculiar behavior of the mirror creature remains unexplained. Only the emergence of a new model which combines the two models which are already in infant's possession, i.e., of other people and of the self seen from the first-person perspective ("it's how the others see me!"), provides the perfect fit to the sensory data. This is the internal model of oneself as seen from the third-person perspective. The task of an infant is to restructure one's internal hypothesis space and form such model. Seen this way, mirror self-recognition becomes an intellectual struggle, requiring a strike of insight reflecting mental reorganization similar to out-of-the-box thinking required to solve other ill-posed problems (Knoblich et al., 1999; Lewis et al., 2018; Pezzulo et al., 2014; Sternberg and Davidson, 1995; Weisberg, 2015).

If mirror self-recognition is just an ill-posed problem then the ability to recognize oneself in a mirror should not rely on an innate cognitive module (Barrett and Kurzban, 2006; Samuels, 2012), but on more general cognitive skills. Therefore, it should be possible to teach it to many animals which are traditionally believed to be incapable of it. However, because these animals may possess weaker cognitive skills, it may require long-term extensive training in a heavily simplified setup. Traditionally it is assumed that among primates only great apes can exhibit mirror self-recognition (Suddendorf and Butler, 2013). However, recent studies found that with appropriately long training it is possible to teach this ability to rhesus monkeys (Chang et al., 2015; Chang et al., 2017). These results suggest that mirror self-recognition may rely on general learning mechanisms rather than cognitive modules acquired through evolution.

The view that mirror self-recognition is acquired through acquisition of a new internal model is also in line with conservative accounts of significance of acquisition of this skill. Rich views of MSR argue that it marks the beginning of self-awareness (Gallup, 1970; Gallup et al., 2014). Conservative views disagree with this and postulate that MSR is a much less profound cognitive skill and provide a number of arguments supporting it (Rochat and Zahavi, 2011; Suddendorf and Butler, 2013). A conservative view is also more compatible with data showing that mirror self-recognition marks only the beginning of the longer process of acquisition of one's third-person visual representation. Toddlers which pass the classical rouge mark test with mirrors typically need several more months until they can do it with live video recordings, and it is only around the age of four when they are able to recognize themselves in a delayed video recording or on a photography (Povinelli et al., 1996; Suddendorf et al., 2007).

Is there evidence that the third-person visual representation of oneself is underpinned by distinct brain areas than other self-representations? Data described in the previous section suggest that it is possible to experience disruptions of the first-person visual representation without accompanied disturbances of the third-person self-representation (Tobita et al., 1995; Verret and Lapresle, 1978). Moreover, instances of mirror agnosia show that lesions of the parietal cortex can lead to selective loss of the ability to comprehend how a mirror works (Binkofski et al., 1999; Ramachandran et al., 1997). Are there cases of people who fail to recognize themselves in a mirror, even though they retain the ability to use mirrors otherwise? Indeed, there are rare reports of mirrored self-misidentification following brain lesions in the right hemisphere (Villarejo et al., 2011; also following hypnotic suggestion: Connors et al., 2012) suggesting that the third-person self-representation may be underpinned by different brain structures than the visual first-person self-model. However, as indicated in the previous section – more research is needed to fully evaluate this possibility.

3.4 Extended, social, and abstract self-representations

Self-representations described so far all reflect different aspects of the self understood as one's body and are grounded in information coming from specific sensory modalities. However, our self extends beyond our bodily self and our senses. Non-bodily and abstract self-representations (group membership, nationality, religion, abstract beliefs about ourselves and our bodies) are standard targets of social and cognitive theories of self and identity, not only in cognitive science, but also in social psychology (Baumeister, 1999; Baumeister and Tice, 1986; Markus, 1977; McCall and Simmons, 1966; Tajfel, 1982), sociology (Goffman, 1956; Mead, 1934; Owens et al., 2010; Stryker, 1968, 2008), political psychology (Jenke and Huettel, 2016; Van Bavel and Pereira, 2018; Whitehouse, 2018) and political science (Hayward and Watson, 2010; Huddy, 2001; Kalin and Sambanis, 2018) to name a few. They can include such diverse entities as representations of one's clothes and possessions, territory, family and friends, group membership, personality traits, episodic memories, and many others (James, 1890).

How can this type of self-representation emerge in development? It may appear that the underlying mechanism must be different from the one involved in development of the bodily self-representations which were described so far. However, I will argue that they can emerge as a consequence of applying the same general rule as postulated for the emergence of other self-representations, i.e., differentiating between self- and non-self-related information. In this case, however, the scope of possible self-relatedness extends beyond one's physical body and beyond sensory information. As a consequence, the mechanism underlying differentiation cannot rely on sensorimotor congruence, because we do not have motoric control over abstract entities and control over extra-bodily physical entities is typically only temporary.

What then, can be the underlying mechanism? The answer advocated here is that humans likely possess innate forms of the extended and social self-models. These two self-models, together with the bodily self-models, may later form the basis for subsequent development of the abstract self-representation through similar processes of formation of new internal models to the one advocated in regard to the bodily selves. It can be summarized through three postulates:

1. Humans possess evolutionarily inbuilt precursors of the extended self-model
2. Humans possess evolutionarily inbuilt precursors of the social self-model
3. Abstract self-model emerges on the basis of the extended, social and bodily self-models

I will discuss each of these claims in the following three sections.

3.4.1 The extended self

The extended self is usually defined as the aspect of the self composed of representations of one's possessions (Belk, 1988, 2013; Kim and Johnson, 2014). At the most fundamental level, representing something as mine (possession over something) reflects a cognitive ability to differentiate between objects in the environment which belong to me and the ones which do not. As such, it reflects the same fundamental classification operation (into "me" and "not-me") as the one described earlier, although in this case it classifies extracorporeal objects and entities.

Human toddlers begin to demonstrate understanding possession at a similar time to when they begin to show signs of mirror self-recognition, usually when they are between 18 and 24-month old (Fasig, 2000; Rochat, 2011; Rodgon and Rashman, 1976). At this stage they begin to use possessive pronouns (e.g., "Mine!"), and exhibit defensive behavior in regard to objects which belong to them. Given these indications, by roughly 2 years of age the majority of toddlers appear to develop the explicit concept of possession, and as a consequence, acquire a basic form of the extended self. The fact that it develops so late might suggest that an internal model of possession is an advanced, high-level ability. Moreover, because it appears only after a toddler begins to speak, it suggests that representing possession may require language and language-based reasoning.

If representing own possession is such an advanced cognitive achievement then it should be absent in non-human animals. Although it has been rarely directly discussed in biology, there are phenomena indicating that certain forms of this capacity might be in fact widespread in animal kingdom (Strassmann and Queller, 2014). One example of representing extra-corporeal entities as self-related is the phenomenon of territoriality. Territoriality reflects a situation in which an animal lives and hunts or forages on a certain territory, but also actively defends it when a different animal or group of animals (typically of the same species) enters it (Noble, 1939). In order to do it, an animal must be able to recognize one's territory and to represent it as "mine," and hence to possess at least some rudimentary "self-territory model." Territoriality is widespread among not only vertebrates, but also invertebrates (Hinsch and Komdeur, 2017; Stamps, 1994) showing that a basic form of the extended self does not require neither language nor high-level cognition. Others examples of behaviors indicating presence of basic forms of extended self are nesting, food caching, and building structures serving to attract potential mates (Ancrenaz et al., 2004; Borgia, 1995; Kaplan, 2015; Plumptre and Reynolds, 1997; Strassmann and Queller, 2014). In all of these cases animals are able to differentiate between non-bodily objects that belong to them and the ones that belong to others, what constitutes the defining characteristic of the extended self.

Within the Bayesian framework the ability to perform discrimination in regard to a territory or a nest suggests that these animals possess internal models (representations) which allow them to classify their perceived environment as either "my" territory (or nest, food, etc.) or not. While such internal model are typically domain-specific and much more restricted than the full-blown human concept of possession, they show that evolutionary precursors of the extended self might be innate also in humans, and the fact that toddlers manifest it only during the second year of life is caused by late maturation of the underlying neural circuitry (see also: Nancekivell et al., 2018).

3.4.2 The social self

Animal behaviors suggest that many of them also possess some forms of a primordial social self. The social self, in line with the definition adopted in this paper, can be understood as underpinned by internal models which allow to distinguish between creatures (typically belonging to the same species) which are in some way related to "me," and the ones which are not. The existence of a form of the social self can be inferred if a human or an animal behaves in a distinctive way toward certain individuals (for example by defending them - typically one's mate, offspring, or members of the same group), as contrasted to its behavior toward other individuals. This ability might be absent in many species. For example, many species of fish breed by spawning a colony of eggs which are abandoned after laying them. In such cases there is no need for mating individuals to be able to identify each other after the mating occurred (and in species with external fertilization even during the process). However, in many other species, especially mammals and birds, one or both parents take care of their offspring. In this case they exhibit a special set of behaviors toward their children, as opposed to any other young animals of the same species. It includes defending and feeding them, and in some cases even forms of teaching (Hoppitt et al., 2008; Thornton and Raihani, 2010). Some species which live in colonies developed complex mechanisms allowing parents to recognize children from sometimes thousands of pups, like in the case of Mexican free-tailed bats which use vocal and olfactory signals to recognize their pups (Balcombe, 1990; McCracken, 1984; McCracken and Gustin, 1991). Similar rule applies to one's mates. Many species form long-lasting monogamous relationships in which animals cooperate in raising offspring for extended periods of time, sometimes for tens of years (e.g., parrots: Kaplan, 2015). In this case often a male provides food and defends only the female that takes care of his offspring. This behavior, however, requires the male to be able to identify his mate and represent her as such. It means that he must possess an internal model of the mate, as opposed to all other female individuals from that species.

In humans a form of the ability to distinguish between close others and strangers is present from birth. Newborn humans are able to distinguish their mother from other human females very early on, demonstrating a basic form of the social self (Bushnell, 2001; Field et al., 1984). At the age of around 6 months infants develop stranger anxiety (Brooker et al., 2013; Waters et al., 1975). They begin to react with signs of distress during encounters with unfamiliar people, even when mother or other caretaker is present. This behavior typically peaks between 6 and 12 months of age and then decreases in intensity. It suggests that at that age infants can distinguish close others and potentially dangerous strangers, a distinction which may foreshadow ingroup/outgroup classification (cf. Dunham et al., 2008), which forms the basis of group-based forms of the social self, such as different forms of group identity.

In summary, comparative evidence shows that some forms of the social self are widespread in the animal kingdom (as evidenced by selective parental care and mate defence). Moreover, developmental studies showing that human newborns can recognize their mothers suggest that a basic form of social self is present at birth in humans as well.[5]

3.4.3 The abstract self

The abstract self includes representational content that goes beyond representations of one's body and objects or agents in the extra-corporeal space. It includes a wide range of content, but generally it is composed of self-related representations encoded in semantic and episodic memory. This is the type of self which has been visualized by the majority of nodes in Figure 1. Although it is composed of more abstract content thea the notions of the self described so far, the criterion used to determine whether a given representational content forms part of the self or not remains the same: a given representation forms a part of one's (abstract) self if it is represented as self-related.

As such, the abstract self forms a category which includes multiple notions of the self present in psychological, sociological, and related literature. It includes one's content of autobiographic memory (Haslam et al., 2011; Klein et al., 2004; Nelson and Fivush, 2004; Wang, 2004), content of semantic autobiographical memory (Martinelli et al., 2013), representations of one's personality traits – sometimes regarded as constituting the "psychological self" (Hu et al., 2016), and all other forms of semantic self-knowledge (Baumeister, 1999; Conway, 2005; Gillihan and Farah, 2005; Kihlstrom et al., 2003; Markus, 1977; Martinelli et al., 2013; Neisser, 1988). It can include linguistic representations of one's emotional state, metacognitive judgments about oneself (including one's cognition), and abstract judgments about one's body. It also involves various types of identities: national, political, religious, gender, cultural - all of them are formed on the basis of socially and culturally transmitted information, and represent abstract socio-cultural constructs. The proposed model allows to link all of these seemingly disparate fields of study by showing that on the cognitive level they all reflect a manifold of representations stored in the semantic and episodic memory and acquired predominantly through mechanisms of social and cultural learning.

How such abstract self-representation can develop? The proposal is that it emerges based on the same mechanisms of new model acquisition as bodily self-models described before, i.e., by forming (and later reforming) an internal model which allows to classify abstract content into self and non-self. The crucial task for an individual is to learn that it can be applied to virtually any representational content, including highly abstract concepts and categories such as personality traits, philosophical ideas, etc. As argued in the previous sections, it is highly likely that humans are born with some basic capacity to represent objects in the environment (via possession - forming the basis of the extended self) and other people (via emotional attachment - underpinning the social self) as self-related. Moreover, very early in development human infants (but also many other animals) are able to distinguish between their body and the external world, at least when perceived from the first-person perspective. It appears safe to assume that before they begin to exhibit signs of linguistic abstract thoughts, young humans already possess a range of self-representations (bodily, social, and probably basic extended). The task for the developing toddler becomes to notice the underlying rule – that we can classify all representations as self versus non-self-related - and apply it to the emerging abstract representations as well.

The first clear indication that abstract self is emerging comes when one begins to use self-related language. It includes first instances of usage of personal and possessive pronouns ("I," "Me," "mine") and being able to generate self-descriptions. In typically developing humans this usually happens between 18 and 24 month of life (Bates, 1990; Fasig, 2000; Levine, 1983; Stipek et al., 1990; Tomasello, 1998). This is the same period of time as when toddlers begin to recognize themselves in a mirror, but also when they begin to show explicit signs of understanding of the concept of possession (Fasig, 2000; Rochat, 2011). It raises the possibility that these three developmental achievements are related. However, a study by Fasig (2000), in which all three were tested found that while use of self-related language and understanding of possession tend to co-occur, mirror self-recognition was unrelated to them (see also: Levine, 1983). It suggests that in humans the emergence of the abstract self may be related to development of explicit understanding of possession. One possibility is that possession serves as a springboard for development of abstract self-representation. One way to investigate this

[5] Of course the basic human social self, as well as the basic extended self are most likely not direct evolutionary descendants of mechanism that led to the development of these basic social and extended selves in many other animals (especially the ones that are evolutionarily distant from us, like birds). It also means that specific mechanisms that underlie them (at the algorithmic and implementational levels) can be very different in humans.

hypothesis is to conduct longitudinal research tracking the developmental trajectory of personal and possessive pronouns individually, as well as self-descriptions (current methods usually group them together, e.g., Stipek et al., 1990) in order to determine which of them appear earlier and which later. If indeed possession is the basis for explicit linguistic self-representation then toddlers should first begin to understand and produce possessive pronouns ("my," "mine") and only later first-person personal pronouns.

Once the basic form of abstract self-representation emerges (as indicated by usage of self-related language) it becomes subject to further development as a result of socio-cultural learning and individual reasoning. This process can be compared to development of understanding of mathematics, which begins with understanding of individual numbers (one, two, three), then the concept of natural numbers, and then integers, rational and complex numbers (Carey, 2009). According to the proposed theory, abstract self-model changes during the course of human development in analogous way. It can be illustrated by the developmental trajectory of self-descriptions. Very young toddlers use only pronouns and own name for self description. However, as they mature their ability to describe themselves rapidly increases. Already before the fourth year of life one's self-description can become very sophisticated, as illustrated by the following example:

> I'm 3 years old, I'm a boy, and my name is Jason. I live with my mommy and daddy who really love me. My mommy makes me yummy spaghetti! I am going to get my own baby sister for Christmas! I have blue eyes and a kitty that is orange and a television in my own room, it's all mine! I know all of my ABC's, listen: A, B, C, D, E, F, G, H, J, L, K, O, P, Q, R, X, Y, Z. I can run real fast, faster than when I was 2. And I can kick a soccer ball real far, all the way from one end of the field to the other. I'm a lot bigger now. When I look in the mirror at me, I can tell I grew. My daddy puts marks on the mirror to show how much taller I get. I have a nice teacher at preschool, she thinks I'm great at everything! I can count up to 100, want to hear me? I can climb to the top of the jungle gym, I'm not scared! I'm never scared! I'm always happy. I'm really strong. I can lift this chair, watch me! My mommy and I like to make up stories about me, she helps me remember things I did or said. Harter (2012, p. 28).

While younger children use mostly concrete concepts referring to external observable characteristics to describe themselves (I'm a girl, I have blond hair, etc.), with time their self-descriptions begin to include more abstract and non-observable concepts (such as psychological traits: Yuill, 1992a, 1992b), and they become more structured and coherent (Damon and Hart, 1991; Harter, 2012). They also tend to take more narrative form, what marks the emergence of autobiographical memory (Klein et al., 2004; Nelson and Fivush, 2004), which forms the basis of what is often regarded as the narrative self (Dennett, 2014; Gallagher, 2000; Schechtman, 2011) or narrative identity (McAdams, 2001; McAdams and McLean, 2013).

The abstract self is a continuously developing collection of semantic knowledge and episodic memories. Importantly, this development is predominantly driven by socio-cultural learning mechanisms, as illustrated by the fact that self-descriptions differ across cultures from very young age (Hart and Edelstein, 1992; Wang, 2004, see also: Nelson and Fivush, 2004). However, the role of socio-cultural environment becomes even more apparent with age. For example, during adolescence, self-descriptions begin to increase in complexity and start to resemble adult ones by increasingly referring to specific sociocultural entities, such us cultural groups (subcultures, social classes), philosophical and social convictions (political and religious beliefs, personal beliefs), preferences (taste in music and movies) etc. As such, one's abstract self becomes increasingly composed of representations of socio-cultural entities, such as nationality, religion, social position, or even spiritual beliefs, rather than the ones rooted in one's body or the physical world. Finally, it may even include metacognitive representations, such as thoughts about one's thought or one's cognition – a situation which is especially vivid in mental disorders. For example in some types of schizophrenia an affected individual may develop beliefs that some of their thoughts belong to another person (Bortolotti and Broome, 2009; Martin and Pacherie, 2013; Young, 2008). In this case one begins to represent some of one's thoughts as self-related, while the others as belonging to someone else.

A fully developed abstract self is specific to humans only, because according to the current knowledge no other animals possess language and means for cultural learning of highly abstract concepts. However, the question to what extent other animals can possess a basic form of the abstract self is difficult to address. It may be possible for an organism to acquire some form of an abstract self-model in the absence of language. If an animal can represent possession, or be able to recognize itself in a mirror, then perhaps it may be also capable to develop a basic form of the abstract self-representation.

4 Innate and learned models

The proposed model assumes that the brain performs approximate Bayesian inference. Under this approach the brain is seen as an inference machine which is composed of (hierarchically organized) internal models of hidden causes of sensory input (Clark, 2016; Hohwy, 2013). The proposed model suggests that there is a neural mechanism which allows new models to emerge as a result of recombination or modification of the existing models (see for example: Lake et al., 2018; Smith et al., 2019). As such, it postulates how new models can be acquired during the lifespan of an individual. However, cognitive representations can also be acquired in phylogeny, i.e., over the time course of multiple generations, as a result of evolution. Indeed, there is strong evidence that humans are born with an evolutionarily hard-wired stock of representations. For example, newborns appear to be able to detect faces (Johnson, 2005; Johnson et al., 1991), and to discriminate biological from non-biological motion (Bardi et al., 2011, 2014; Bottari et al., 2015). Moreover, a rudimentary preference for face-like stimuli can be detected in foetuses even before birth (Reid et al., 2017). These results strongly suggest that newborns possess basic forms of internal models of faces and biological motion. Because visual experience in the uterus is extremely limited, these representations could not have been acquired via mechanisms of learning. Instead, they must have been genetically encoded.

Genetic effects can be easily accommodated into Bayesian models of cognition in the form of innate models. For example, in predictive coding and related theory of free energy principle evolutionary effects are understood as instances in which models of hidden causes are obtained through evolutionary processes (Sims, 2017) and understood as "optimizing the agent's model and priors through neurodevelopment and natural selection" (Friston, 2010). As such, the problem of innateness can be rephrased as the problem of which models are acquired through evolution, and which need to be learned.

The proposal introduced in this paper aimed for maximal parsimony and therefore postulated innate models only where it seemed unlikely that learning plays a decisive role. These exceptions involve primitive forms of extended and social self which can be observed in multiple animal species, as manifested through, e.g., territorial behavior and defence of a mate or offspring (Figure 7). Moreover, these behaviors are manifested universally (i.e., often in almost all animals of a given species) and are highly stereotypical – which are characteristics of innately specified mechanisms (Gross and Rey, 2012). However, it is also possible that in humans they are learned – especially in the case of the extended self which becomes evident only around the second year of life.

It is also possible that some self-models which in principle could be learned, are inborn. Biological organisms can greatly benefit from having an innate stock of mechanisms and representations, including some types of internal models of the self. The proposed primordial self-other distinction is almost certainly innate in humans. One line of evidence comes from reports of phantom limb sensations in people with congenital absence of limbs, a condition which is also known as aplasic phantoms (Brugger et al., 2000; Gallagher et al., 1998; Melzack et al., 1997). Because such individuals do not possess limbs it was not possible for them to acquire parts of their sensori-motor self representing their missing limbs through learning. The fact that they can nevertheless experience phantom limbs suggests that these parts of self-representation must be genetically pre-programmed. However, there are alternative explanations of this phenomenon, which do not need one to postulate the existence of innate representations (Gallagher, 2006, chapter 4; Blumberg and Dooley, 2017). At present this debate is still unresolved.

At the same time, there are stages of development of the self which almost certainly are acquired through individual and cultural learning. There are strong reasons to believe that it includes mirror self-recognition and acquisition of abstract self-representation. In these cases mechanisms of new model acquisition proposed above could illustrate their full power. This might be especially true in regard to various high-level components of the abstract self-representation, such as national and religious identity, where various forms of cultural learning may be critical (Heyes, 2018; Mesoudi, 2011; Tomasello et al., 1993).

5 Empirical predictions and limitations

Bayesian approaches to cognition are sometimes criticized as being unfalsifiable (Bowers and Davis, 2012a, 2012b; Jones and Love, 2011), what raised a considerable discussion. Here I will follow the answer to it provided by Griffiths et al. (2012):

> In evaluating claims about falsifiability, it is useful to distinguish between a model and a theoretical framework. A model is proposed to account for a specific phenomenon and makes specific assumptions in order to do so. A theoretical framework provides a general perspective and a set of tools for making models […] Models are falsifiable, but frameworks are typically not. Rather, frameworks live or die based on their ability to generate models that are useful. […] We believe that specific Bayesian models are readily falsifiable (or, at least, as falsifiable as any empirical hypothesis—any hypothesis can be "saved" by suitable *ad hoc* adjustments to other aspects of the theory […]).

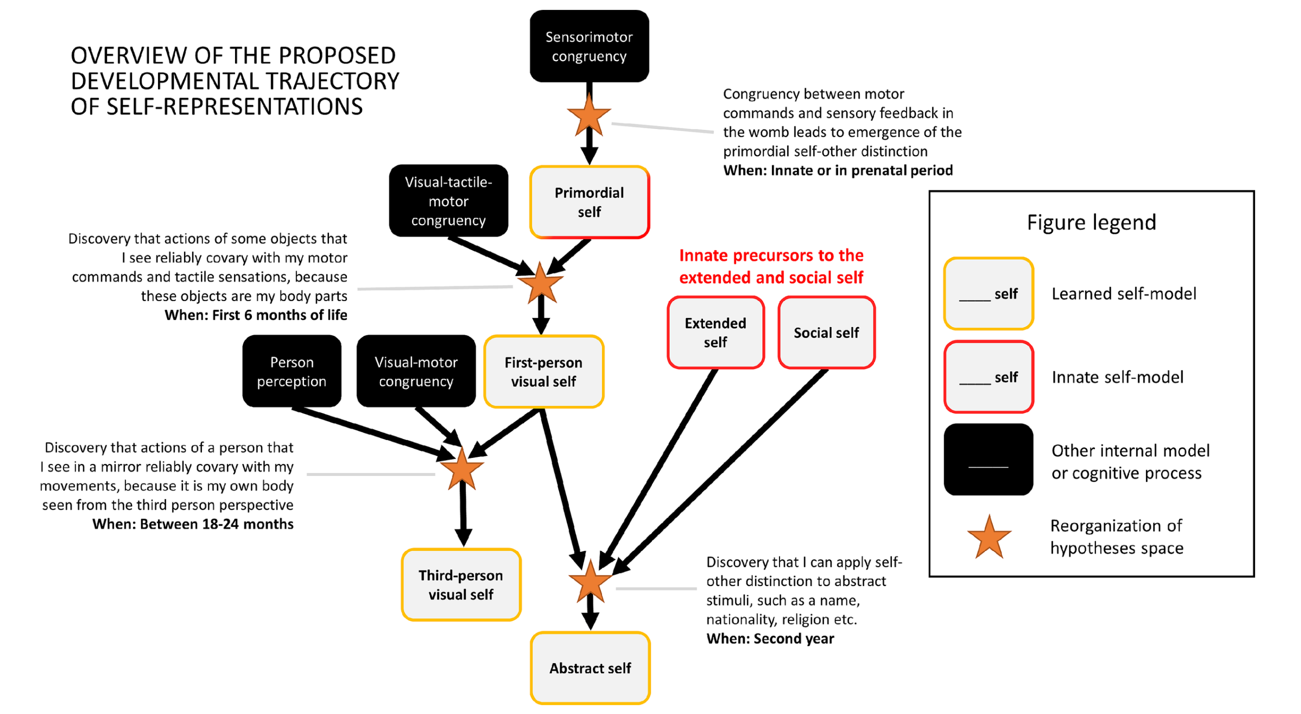
OVERVIEW OF THE PROPOSED DEVELOPMENTAL TRAJECTORY OF SELF-REPRESENTATIONS
Sensorimotor congruency
Congruency between motor commands and sensory feedback in the womb leads to emergence of the primordial self-other distinction
When: Innate or in prenatal period
Visual-tactile-motor congruency
Primordial self
Discovery that actions of some objects that I see reliably covary with my motor commands and tactile sensations, because these objects are my body parts
When: First 6 months of life
Innate precursors to the extended and social self
Extended self
Social self
Person perception
Visual-motor congruency
First-person visual self
Discovery that actions of a person that I see in a mirror reliably covary with my movements, because it is my own body seen from the third person perspective
When: Between 18-24 months
Third-person visual self
Discovery that I can apply self-other distinction to abstract stimuli, such as a name, nationality, religion etc.
When: Second year
Abstract self
Figure legend
____ self
Learned self-model
____ self
Innate self-model
____
Other internal model or cognitive process
Reorganization of hypotheses space

FIGURE 7
Overview of proposed stages of development of self-representations.

TABLE 1 Examples of actual and potential evidence for double dissociations between the postulated self-models.

Self-representation	Evidence of presence	Evidence (potential or existing) of absence, loss or disruption
Primordial (sensorimotor) self	Being able to distinguish between own body and the external world (objects, other agents) based on touch and proprioception	Clinical cases of loss of sense of body ownership: alien hand syndrome, anarchic hand syndrome, certain cases of somatoparaphrenia
First-person visual self	Successful recognition of one's body parts when they are seen from the first-person perspective	Cases of selective disruption (Tobita et al., 1995; Verret and Lapresle, 1978). It is probably absent in congenitally blind individuals
Third-person visual self	Successful self-recognition in a mirror, an image or a video	Some cases of mirror agnosia. Developmentally absent in humans before 18 months. Absent in most non-human animals.
Extended self	Territorial behaviour, nesting, defence of one's possessions, verbal reports of possession	It should be possible to observe loss of the ability to differentiate between self and other-owned objects, territory etc.
Social self	The capacity to preferentially treat one's mate, offspring or group members.	It should be possible to observe loss of the ability to distinguish between close others and strangers (potentially in Capgras and Fregoli delusions)
Abstract self	The capacity to speak about abstract concepts and entities as being self-related	It should be possible to observe selective loss of the ability to use first-person personal pronouns. Abstract self is absent in animals and humans that do not possess the capacity for abstract thought or language.

> But the general Bayesian approach, as with any scientific framework, is not. Griffiths et al. (2012) p. 416.

The proposal of developmental trajectory of self-representations described in section 3 is a specific model that can be readily tested and the subsequent subsection will discuss how to do it. The general Bayesian framework is not falsifiable and it should be rather evaluated on the criteria of how useful for generating models it is. However, the current paper discusses not only a specific model of the developmental trajectory, but also a more general process of how development can take place in a Bayesian brain, which can be also treated as a smaller-scale framework. I will discuss it in the subsequent subsection.

5.1 How to test the model of development of self-representations

The proposal outlined in section 3 of this paper yields specific empirical predictions: if our self-representation is underpinned by multiple distinct internal models then we should observe evidence of double dissociations between these models. In principle, there should exist situations in which each individual postulated self-model is selectively present or absent. Phylogeny and ontogeny show examples in which more basic models are present, while more advanced models are absent, providing evidence for one part of this dissociation. An example is the dissociation between the first-person (1PVSR) and the third-person visual self-recognition (3PVSR). While 1PVSR is widespread in animal kingdom, the 3PVSR (as measured by mirror self-recognition) is not. Importantly, the same individuals (infants or animals) that at an earlier time point are not able to recognize themselves in a mirror, at a later point begin to display this ability. This illustrates that it is possible to possess either 1PVSR alone or to possess both. However, a complete proof of a double dissociation requires also evidence of a situation in which a developmentally earlier model is lost, while the later model is retained.[6] This would rule out the possibility that the same model is responsible for both 1PVSR and 3PVSR, and that the emergence of mirror self-recognition only marks a developmental change in the internal model responsible for 1PVSR. While fully conclusive evidence is missing, there are neurological case studies (described in section 3.2 and 3.3: Fotopoulou et al., 2011; Tobita et al., 1995; Verret and Lapresle, 1978) that suggest that these two types of self-recognition are underpinned by two distinct models. However, more detailed analysis of the existing cases, as well as more research, are needed to validate this issue.

The same approach should be applied to each other pair of postulated models. The biggest challenge for the proposed theory lies in providing evidence of selective loss of earlier models, with preserved later models. This would mean, for example, demonstrating selective loss of aspects of the social or extended self, with otherwise intact abstract self-representation, or a selective loss of the primordial (sensorimotor) self-representation. Table 1 lists examples of evidence that would validate the proposed theory.

5.2 Can we test the proposed developmental mechanism?

It is possible that the general logic of the proposed theory is correct, i.e., that self-representation emerges in a series of discrete steps through acquisition of new self-models, but that the specific trajectory from section 3 is wrong in respect to certain specific aspects. For example, it might turn out that the emergent models are not single, individual entities, but collections of several models. For example, it is possible that

[6] Because the developmentally later models are constructed based on the earlier models, for such dissociation to be observed the loss of an earlier model (for example due to a brain lesion) can occur only after the later-model has been already acquired.

adult humans possess not one, but several distinct sensorimotor self-models, which might be composed of multiple models responsible for each half of the body or each body part. Similarly, humans might possess several third-person visual self-representations (one responsible for mirrors, another for video recordings, etc.). Further, it is similarly likely that there is not one model for abstract self-representation, but a collection of them, with each responsible for a different aspect of abstract self-representation. In a similar vein, there might be other independent self-models that do not fit neatly into the proposed classification, such as the agentive self proposed by Riva (2018).[7] As such, the specific model proposed here will most likely require updating in the light of newly collected data. This, however, brings back the issue of falsifiability.

Understood in this way, the more general proposal introduced in this paper can be treated as a framework, rather than as a specific model. As argued by Griffiths et al. (2012) frameworks are generally not falsifiable. However, there are situations in which the proposed developmental mechanism can become validated. These would involve demonstrating that the basic assumptions of the framework are incorrect:

A. The framework assumes that there is more than one self-model. If we observe that there is a single self-model that is responsible for all types of self-representations then this would falsify it.
B. It assumes that at least some self-models emerge through learning. Demonstrating that all self-models (or self-representations) are innate would falsify the framework. Moreover, the framework is most useful if all self-models are learned. Conversely, the more self-models are shown to be innate the lower the explanatory power of the proposed framework.
C. It assumes that we acquire new self-models through learning rather than maturation. Evidence that an organism possesses precursors to some self-model that gets activated by specific triggers would reduce the explanatory power of the framework. The most famous example of such mechanism is the phenomenon of imprinting (Hess, 1959; Lorenz, 1935), but similar mechanisms have been discussed for development of mirror self-recognition (Anderson and Gallup, 2015; Suddendorf and Butler, 2013).
D. It assumes that we generally acquire new self-models through a specific form of learning, i.e., by recombining existing knowledge into new internal models.[8] It means that evidence that we acquire new self-model only through different forms of learning could falsify the model.

Overall, the proposed framework describes acquisition of new self-models as a process that critically depends on a very specific form of learning, so any evidence of innateness, maturation, or alternative learning mechanisms directly undermine it.

6 Comparison with other theories

The proposed model of development of self-representation bears many similarities to the model recently proposed by Giuseppe Riva (2018). Both of these models postulate that the self emerges in a series of discrete developmental steps that lead to clearly distinguishable representations. Riva proposes six such representations: (1) the innate Sentient Body, which is an invariant spatial structure that integrates interoceptive, proprioceptive and vestibular signals and underpins the minimal phenomenal selfhood, (2) the Spatial Body that develops in the first 6 months of life and underpins self-location, (3) the Active Body that develops from the second half of the first year of life and underpins one's sense of agency, (4) the Personal Body, forming one's whole-body representation, which is responsible for the first-person reflective experience of owning a whole body (5) the Objectified Body, which stands for one's third-person body self-representation and (6) the Social Body which integrates body-related social rules and narratives and underpins one's sense of body satisfaction.

Many of these representations overlap with the current proposal: the primordial self is reminiscent of Riva's Sentient Body and the postulated third-person visual self-representation corresponds to his Objectified Body. Moreover, Riva's Social Body in the current proposal forms an important part of the abstract self-representation. However, in respect to the other developmental steps the two theories make different proposals. This comes as a consequence of differences in perspective: the current proposal differentiates self-representations (self-models) primarily based on types of information that underpin them: visual, auditory, tactile, abstract. On the other hand, Riva distinguishes body representations primarily based on the function that they play in constituting specific aspects of the self, i.e.: minimal phenomenal experience, self-location, agency, sense of whole body ownership, capacity for third-person self-reflection, and body satisfaction. This difference of perspective leads to important differences in how many and what types of self-representations one should postulate.

Other recent theories focused specifically on highlighting the role of social contact in development of the self. Humans spend the first 9 months of their development within the body of another person – their mother. It means that newly conceived humans are in direct social contact with another person from day one (even before they develop a neural system). This is a profoundly important insight and several recent papers put emphasis on it (Ciaunica, Constant, et al., 2021; Ciaunica, Safron, et al., 2021). While acknowledging its importance, the proposal outlined in this paper does not predict that this fact should make a fundamental difference to the emergence of

[7] The model described in section 3 focused on self-representations that are relatively stable across time, as this is how this term is usually understood. However, it remains an open question whether it can be extended to describe how we represent our actions (what would underpin our sense of agency). In such case one would need to either postulate existence of an additional self-model or argue that it can be accounted by the already postulated self-models. More on sense of agency: Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*, *18*(4), 196.

[8] An additional challenge is how to prove that new models can or have been acquired through creative recombination of the existing models. This issue is investigated within the area of study of cognitive processes underpinning creativity. However, it would require a more lengthy discussion that goes beyond the scope of this paper. See e.g.: Kaufman, J. C., & Sternberg, R. J. (2010). *The Cambridge handbook of creativity*. Cambridge University Press.

the primordial self-model. In the postulated developmental mechanism outlined in section 3.1 the contrast that drives the emergence of the primordial self-model is between sensory stimulation that can be predicted by one's motor commands versus stimulation that cannot be predicted. From this perspective the womb and the rest of mother's body are likely treated as elements of the external world. However, it is possible that due to tight mechanical and physiological coupling with mother's body parts of it can become represented by a foetus' nervous system as part of oneself, and that this classification has important consequences in postnatal life.

Another important aspect raised in other recent theories is the role of sociality, and especially social touch, in development of self-representation. This topic has been recently extensively discussed in several papers (Atzil et al., 2018; Ciaunica and Fotopoulou, 2017; Fotopoulou and Tsakiris, 2017). In the current proposal social interaction is important in regard to the social, abstract and extended self: a person who is deprived of any social contact would likely fail to develop the social and extended self-models beyond what is genetically hard-wired. Moreover, the abstract self-model, which is fundamentally reliant on social and cultural learning, would be absent altogether. However, the current proposal does not treat social affiliative contact or social touch as necessary factors for development of any of the proposed self-models. It means that people who were in a social environment, but were deprived of physical contact with others (like children raised in orphanages in Romania under Nicolae Ceaușescu's regime) should still be able to develop all of the postulated self-models, although deprivation of sociality and social touch, especially in childhood, might strongly affect how quickly and in what form each model develops. This might naturally have great impact on one's social self, leading socially deprived people to develop altered self-representations (Carlson and Earls, 1997; Nelson III et al., 2007).

7 Conclusion

The current paper proposed a novel theory of the self understood as a representational structure in a Bayesian brain. By taking a developmental perspective it linked research traditions on bodily, abstract, social and extended self and proposed of a framework in which different facets of the self are understood as different stages of development of one's internal self-models. The goal was, following the approach of William James (1890), to provide a comprehensive classification of different types of the self, but also to suggest how they may be related to each other. The theory advocated in this paper yields testable predictions regarding the structure of self-representation not only in humans but also in other animals. It allows to describe and explain selective losses of components of the self (e.g., first-person visual self-representation, third-person visual self-representation) and situate them within one general framework.

The proposed theory may be also regarded as a case study of how the process of acquisition of new internal models can be understood in the context of Bayesian models of cognition. It illustrates how using this framework may be beneficial. First, it provides a unifying account of diverse empirical phenomena, including (in the case of the self) bodily self-representation, social identity, territoriality, possession, and abstract self-related thoughts. Second, it suggests testable predictions which allow to test diverse models in order to determine the architecture underlying self-representations. In the interest of brevity this paper focused mainly on case studies of brain lesion showing dissociations between different types of self-models, as they provide the strongest evidence for specific selective impairments. However, other types of empirical evidence, such as neuroimaging and electrophysiology, can be equally important in further investigations of the structure of the self.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

MW: Conceptualization, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Writing of this manuscript was done while the author was employed in the following grants: the European Research Council under the European Union's Horizon 2020 research and innovation program, ERC Starting grant ERC-2016-StG-715058, awarded to Agnieszka Wykowska, the Australian Research Council grant DP160102770 awarded to Jakob Hohwy, and the European Research Council under the European Union's Seventh Framework Program (FP7/2007–2013) / ERC grant agreement n° 609819, Constructing Social Minds: Coordination, Communication, and Cultural Transmission (SOMICS) awarded to Guenther Knoblich and others.

Acknowledgments

The ideas discussed in this manuscript were developed and discussed during the time that I spent at the Social Mind & Body Lab led by Guenther Knoblich & Natalie Sebanz at Central European University, Cognition & Philosophy Lab led by Jakob Hohwy at Monash University and the Social Cognition in Human-Robot Interaction Group led by Agnieszka Wykowska at Italian Institute of Technology. They were also inspired by the content of two courses at Central European University: "The origin of concepts" by Gergely Csibra & György Gergely, and "Visual perception and learning in the brain" by Jozsef Fiser. I also want to thank Piotr Wolski and Michał Wierzchoń from Jagiellonian University in Kraków for comments on my dissertation in which I described early version of these ideas. Finally, I want to thank the reviewers for constructive criticism and many thought provoking questions that greatly helped to improve the manuscript.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Allen, M., and Tsakiris, M. (2018). The body as first prior: Interoceptive predictive processing and the primacy of self-models. In: *The interoceptive mind. From homeostasis to awareness*. eds. M. Tsakiris and De Preester H. (Oxford, UK: Oxford University Press), 27–45.

Amsterdam, B. (1972). Mirror self-image reactions before age two. *Develop. Psychobiol.* 5, 297–305. doi: 10.1002/dev.420050403

Ancrenaz, M., Calaque, R., and Lackman-Ancrenaz, I. (2004). Orangutan nesting behavior in disturbed forest of Sabah, Malaysia: implications for nest census. *Int. J. Primatol.* 25, 983–1000. doi: 10.1023/B:IJOP.0000043347.84757.9a

Anderson, J. R., and Gallup, G. G. (2015). Mirror self-recognition: a review and critique of attempts to promote and engineer self-recognition in primates. *Primates* 56, 317–326. doi: 10.1007/s10329-015-0488-9

Anderson, A. L., and Thomason, M. E. (2013). Functional plasticity before the cradle: a review of neural functional imaging in the human fetus. *Neurosci. Biobehav. Rev.* 37, 2220–2232. doi: 10.1016/j.neubiorev.2013.03.013

Apps, M. A., and Tsakiris, M. (2014). The free-energy self: a predictive coding account of self-recognition. *Neurosci. Biobehav. Rev.* 41, 85–97. doi: 10.1016/j.neubiorev.2013.01.029

Arduini, D., Rizzo, G., and Romanini, C. (2013). Fetal behavioral states and behavioral transitions in normal and compromised fetuses. In: *Fetal Development*. eds. J. Lecanuet, W. P. Fifer, N. A. Krasnegor and W. P. Smotherman (New York. USA: Psychology Press), pp. 83–99.

Atzil, S., Gao, W., Fradkin, I., and Barrett, L. F. (2018). Growing a social brain. *Nat. Hum. Behav.* 2, 624–636. doi: 10.1038/s41562-018-0384-6

Babo-Rebelo, M., and Tallon-Baudry, C. (2018). Interoceptive signals, brain dynamics, and subjectivity. In: *The interoceptive mind. From homeostasis to awareness* eds. M. Tsakiris and De Preester H. (Oxford, UK: Oxford University Press), p. 46–62.

Baker, B., Lansdell, B., and Kording, K. P. (2022). Three aspects of representation in neuroscience. *Trends Cogn. Sci.* 26, 942–958. doi: 10.1016/j.tics.2022.08.014

Balcombe, J. P. (1990). Vocal recognition of pups by mother Mexican free-tailed bats, Tadarida brasiliensis mexicana. *Anim. Behav.* 39, 960–966. doi: 10.1016/S0003-3472(05)80961-3

Bardi, L., Regolin, L., and Simion, F. (2011). Biological motion preference in humans at birth: role of dynamic and configural properties. *Dev. Sci.* 14, 353–359. doi: 10.1111/j.1467-7687.2010.00985.x

Bardi, L., Regolin, L., and Simion, F. (2014). The first time ever I saw your feet: inversion effect in newborns' sensitivity to biological motion. *Dev. Psychol.* 50, 986–993. doi: 10.1037/a0034678

Barrett, H. C., and Kurzban, R. (2006). Modularity in cognition: framing the debate. *Psychol. Rev.* 113, 628–647. doi: 10.1037/0033-295X.113.3.628

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038

Bates, E. (1990). Language about me and you: Pronominal reference and the emerging concept of self. in *The self in transition: Infancy to childhood*. eds. D. Cicchetti and M. Beeghly (Chicago, USA: University of Chicago Press), pp. 165–182.

Baumeister, R. F. (1999). The self in social psychology. London, UK: Psychology Press.

Baumeister, R. F., and Tice, D. M. (1986). "Four selves, two motives, and a substitute process self-regulation model" in *Public self and private self*. ed. R. E. Baumeister (New York, USA: Springer), 63–74.

Belk, R. W. (1988). Possessions and the extended self. *J. Consum. Res.* 15, 139–168. doi: 10.1086/209154

Belk, R. W. (2013). Extended self in a digital world. *J. Consum. Res.* 40, 477–500. doi: 10.1086/671052

Bertenthal, B. I., and Fischer, K. W. (1978). Development of self-recognition in the infant. *Dev. Psychol.* 14, 44–50. doi: 10.1037/0012-1649.14.1.44

Binkofski, F., Buccino, G., Dohle, C., Seitz, R., and Freund, H. J. (1999). Mirror agnosia and mirror ataxia constitute different parietal lobe disorders. *Ann. Neurol.* 46, 51–61. doi: 10.1002/1531-8249(199907)46:1<51::AID-ANA9>3.0.CO;2-Q

Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. *Nat. Rev. Neurosci.* 13, 556–571. doi: 10.1038/nrn3292

Blanke, O., and Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends Cogn. Sci.* 13, 7–13. doi: 10.1016/j.tics.2008.10.003

Blumberg, M. S., and Dooley, J. C. (2017). Phantom limbs, neuroprosthetics, and the developmental origins of embodiment. *Trends Neurosci.* 40, 603–612. doi: 10.1016/j.tins.2017.07.003

Bolis, D., and Schilbach, L. (2020). 'I interact therefore I am': the self as a historical product of dialectical attunement. *Topoi* 39, 521–534. doi: 10.1007/s11245-018-9574-0

Borgia, G. (1995). Complex male display and female choice in the spotted bowerbird: specialized functions for different bower decorations. *Anim. Behav.* 49, 1291–1301. doi: 10.1006/anbe.1995.0161

Bortolotti, L., and Broome, M. (2009). A role for ownership and authorship in the analysis of thought insertion. *Phenomenol. Cogn. Sci.* 8, 205–224. doi: 10.1007/s11097-008-9109-z

Bottari, D., Troje, N. F., Ley, P., Hense, M., Kekunnaya, R., and Röder, B. (2015). The neural development of the biological motion processing system does not rely on early visual input. *Cortex* 71, 359–367. doi: 10.1016/j.cortex.2015.07.029

Bowers, J. S., and Davis, C. J. (2012a). Bayesian just-so stories in psychology and neuroscience. *Psychol. Bull.* 138, 389–414. doi: 10.1037/a0026450

Bowers, J. S., and Davis, C. J. (2012b). Is that what Bayesians believe? Reply to Griffiths, Chater, Norris, and Pouget. *Psychol. Bull.* 138, 423–426. doi: 10.1037/a0027750

Bowman, C. R., and Zeithamova, D. (2018). Abstract memory representations in the ventromedial prefrontal cortex and hippocampus support concept generalization. *J. Neurosci.* 38, 2605–2614. doi: 10.1523/JNEUROSCI.2811-17.2018

Brooker, R. J., Buss, K. A., Lemery-Chalfant, K., Aksan, N., Davidson, R. J., and Goldsmith, H. H. (2013). The development of stranger fear in infancy and toddlerhood: normative development, individual differences, antecedents, and outcomes. *Dev. Sci.* 16, 864–878. doi: 10.1111/desc.12058

Brugger, P., Kollias, S. S., Müri, R. M., Crelier, G., Hepp-Reymond, M.-C., and Regard, M. (2000). Beyond re-membering: phantom sensations of congenitally absent limbs. *Proc. Natl. Acad. Sci.* 97, 6167–6172. doi: 10.1073/pnas.100510697

Bushnell, I. (2001). Mother's face recognition in newborn infants: learning and memory. *Infant. Child Dev.* 10, 67–74. doi: 10.1002/icd.248

Carey, S. (2009). The origin of concepts. Oxford, UK: Oxford University Press.

Carlson, M., and Earls, F. (1997). Psychological and neuroendocrinological sequelae of early social deprivation in institutionalized children in Romania. *The integrative neurobiology of affiliation*. eds. C. S. Carter, I. I. Lederhendler and B. Kirkpatrick (New York, USA: New York Academy of Sciences), 807, 419–428.

Castiello, U., Becchio, C., Zoia, S., Nelini, C., Sartori, L., Blason, L., et al. (2010). Wired to be social: the ontogeny of human interaction. *PLoS One* 5:e13199. doi: 10.1371/journal.pone.0013199

Chalmers, D. J. (2004). "The representational character of experience" in The future for philosophy. ed. B. Leiter (Oxford, UK: Oxford University Press), 153–181.

Chang, L., Fang, Q., Zhang, S., Poo, M., and Gong, N. (2015). Mirror-induced self-directed behaviors in rhesus monkeys after visual-somatosensory training. *Curr. Biol.* 25, 212–217. doi: 10.1016/j.cub.2014.11.016

Chang, L., Zhang, S., Poo, M.-M., and Gong, N. (2017). Spontaneous expression of mirror self-recognition in monkeys after learning precise visual-proprioceptive association for mirror images. *Proc. Natl. Acad. Sci.* 114, 3258–3263. doi: 10.1073/pnas.1620764114

Chen, L., Lambon Ralph, M. A., and Rogers, T. T. (2017). A unified model of human semantic knowledge and its disorders. *Nat. Hum. Behav.* 1:0039. doi: 10.1038/s41562-016-0039

Christoff, K., Cosmelli, D., Legrand, D., and Thompson, E. (2011). Specifying the self for cognitive neuroscience. *Trends Cogn. Sci.* 15, 104–112. doi: 10.1016/j.tics.2011.01.001

Ciaunica, A., Constant, A., Preissl, H., and Fotopoulou, A. (2021a). The first prior: from co-embodiment to co-homeostasis in early life. *Conscious. Cogn.* 91:103117. doi: 10.1016/j.concog.2021.103117

Ciaunica, A., and Crucianelli, L. (2019). Minimal self-awareness: from within a developmental perspective. *J. Conscious. Stud.* 26, 207–226.

Ciaunica, A., and Fotopoulou, A. (2017). The touched self: psychological and philosophical perspectives on proximal intersubjectivity and the self. In: *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*. eds. C. Durt, C. Fuchs and C. Tewes (Cambridge, USA: MIT Press.

Ciaunica, A., Safron, A., and Delafield-Butt, J. (2021b). Back to square one: the bodily roots of conscious experiences in early life. *Neurosci. Conscious.* 2021:niab037.

Clark, A. (2016). Surfing uncertainty: Prediction, action, and the embodied mind. Oxford, UK: Oxford University Press.

Clarke, A., and Tyler, L. K. (2015). Understanding what we see: how we derive meaning from vision. *Trends Cogn. Sci.* 19, 677–687. doi: 10.1016/j.tics.2015.08.008

Connors, M. H., Cox, R. E., Barnier, A. J., Langdon, R., and Coltheart, M. (2012). Mirror agnosia and the mirrored-self misidentification delusion: a hypnotic analogue. *Cogn. Neuropsychiatry* 17, 197–226. doi: 10.1080/13546805.2011.582770

Conway, M. A. (2005). Memory and the self. *J. Mem. Lang.* 53, 594–628. doi: 10.1016/j.jml.2005.08.005

Crane, T. (2003). "The intentional structure of consciousness" in Consciousness: New philosophical perspectives. eds. A. Jokic and Q. Smith (Oxford, UK: Oxford University Press).

Damasio, A. (1999). The feeling of what happens: Body and emotion in the making of consciousness. Boston, USA: Houghton Mifflin Harcour.

Damon, W., and Hart, D. (1991). Self-understanding in childhood and adolescence. Cambridge, UK: Cambridge University Press Archive.

de Klerk, C. C., Filippetti, M. L., and Rigato, S. (2021). The development of body representations: an associative learning account. *Proc. R. Soc. B* 288:20210070. doi: 10.1098/rspb.2021.0070

De Vignemont, F. (2010). Body schema and body image—pros and cons. *Neuropsychologia* 48, 669–680. doi: 10.1016/j.neuropsychologia.2009.09.022

DeCasper, A. J., and Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science* 208, 1174–1176. doi: 10.1126/science.7375928

DeCasper, A. J., Lecanuet, J.-P., Busnel, M.-C., Granier-Deferre, C., and Maugeais, R. (1994). Fetal reactions to recurrent maternal speech. *Infant Behav. Dev.* 17, 159–164. doi: 10.1016/0163-6383(94)90051-5

Dennett, D. C. (2014). "The self as the center of narrative gravity" in *Self and consciousness: Multiple perspectives*. eds. F. Kessel, P. Cole and D. Johnson (London, UK: Psychology Press), pp. 103–115.

DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434. doi: 10.1016/j.neuron.2012.01.010

DiPietro, J., Costigan, K., Shupe, A., Pressman, E., and Johnson, T. (1998). Fetal neurobehavioral development: associations with socioeconomic class and fetal sex. *Dev. Psychobiol.* 33, 79–91. doi: 10.1002/(SICI)1098-2302(199807)33:1<79::AID-DEV7>3.0.CO;2-P

Drysdale, R., and Tsakiris, M. (2021). Growing up a self: On the relation between body image and the experience of the interoceptive body. Body Schema and Body Image: New Directions. Oxford. UK, 210.

Dunham, Y., Baron, A. S., and Banaji, M. R. (2008). The development of implicit intergroup cognition. *Trends Cogn. Sci.* 12, 248–253. doi: 10.1016/j.tics.2008.04.006

Egan, F. (2012). "Representationalism" in The Oxford handbook of philosophy of cognitive science. eds. E. Margolis, R. Samuels and S. Stich (Oxford, UK: Oxford University Press), 250–272.

Eswaran, H., Lowery, C. L., Wilson, J. D., Murphy, P., and Preissl, H. (2004). Functional development of the visual system in human fetus using magnetoencephalography. *Exp. Neurol.* 190, 52–58. doi: 10.1016/j.expneurol.2004.04.007

Fasig, L. G. (2000). Toddlers' understanding of ownership: implications for self-concept development. *Soc. Dev.* 9, 370–382. doi: 10.1111/1467-9507.00131

Field, T. M., Cohen, D., Garcia, R., and Greenberg, R. (1984). Mother-stranger face discrimination by the newborn. *Infant Behav. Dev.* 7, 19–25. doi: 10.1016/S0163-6383(84)80019-3

Filippetti, M. L., Johnson, M. H., Lloyd-Fox, S., Dragovic, D., and Farroni, T. (2013). Body perception in newborns. *Curr. Biol.* 23, 2413–2416. doi: 10.1016/j.cub.2013.10.017

Filippetti, M. L., Orioli, G., Johnson, M. H., and Farroni, T. (2015). Newborn body perception: sensitivity to spatial congruency. *Infancy* 20, 455–465. doi: 10.1111/infa.12083

Filippetti, M. L., and Tsakiris, M. (2018). Just before I recognize myself: the role of Featural and multisensory cues leading up to explicit Mirror self-recognition. *Infancy* 23, 577–590. doi: 10.1111/infa.12236

Fiser, J., Berkes, P., Orbán, G., and Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn. Sci.* 14, 119–130. doi: 10.1016/j.tics.2010.01.003

Fotopoulou, A. (2012). "Towards psychodynamic neuroscience" in From the couch to the lab: Trends in psychodynamic neuroscience. eds. A. Fotopoulou, M. Conway and D. Pfaff (Oxford, UK: Oxford University Press), 25–47.

Fotopoulou, A., Jenkinson, P. M., Tsakiris, M., Haggard, P., Rudd, A., and Kopelman, M. D. (2011). Mirror-view reverses somatoparaphrenia: dissociation between first-and third-person perspectives on body ownership. *Neuropsychologia* 49, 3946–3955. doi: 10.1016/j.neuropsychologia.2011.10.011

Fotopoulou, A., and Tsakiris, M. (2017). Mentalizing homeostasis: the social origins of interoceptive inference. *Neuropsychoanalysis* 19, 3–28. doi: 10.1080/15294145.2017.1294031

Freud, S. (1933). New introductory lectures on psycho-analysis. New York, USA: WW Norton & Company.

Friston, K. J. (2005). A theory of cortical responses. *Phil. Trans. Royal Society London B* 360, 815–836. doi: 10.1098/rstb.2005.1622

Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Friston, K. J., Daunizeau, J., Kilner, J., and Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biol. Cybern.* 102, 227–260. doi: 10.1007/s00422-010-0364-z

Friston, K. J., and Frith, C. (2015). A duet for one. *Conscious. Cogn.* 36, 390–405. doi: 10.1016/j.concog.2014.12.003

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends Cogn. Sci.* 4, 14–21. doi: 10.1016/S1364-6613(99)01417-5

Gallagher, S. (2006). How the body shapes the mind. Oxford, UK: Oxford University Press.

Gallagher, S., Butterworth, G. E., Lew, A., and Cole, J. (1998). Hand–mouth coordination, congenital absence of limb, and evidence for innate body schemas. *Brain Cogn.* 38, 53–65. doi: 10.1006/brcg.1998.1020

Gallup, G. G. (1970). Chimpanzees: self-recognition. *Science* 167, 86–87. doi: 10.1126/science.167.3914.86

Gallup, G. G., Platek, S. M., and Spaulding, K. N. (2014). The nature of visual self-recognition revisited. *Trends Cogn. Sci.* 18, 57–58. doi: 10.1016/j.tics.2013.10.012

Gillihan, S. J., and Farah, M. J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychol. Bull.* 131, 76–97. doi: 10.1037/0033-2909.131.1.76

Goffman, E. (1956). The presentation of self in everyday life. Scotland: Doubleday.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., and Tenenbaum, J. B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. *Trends Cogn. Sci.* 14, 357–364. doi: 10.1016/j.tics.2010.05.004

Griffiths, T. L., Chater, N., Norris, D., and Pouget, A. (2012). How the Bayesians got their beliefs (and what those beliefs actually are): Comment on Bowers and Davis. *Psychol. Bull.* 138, 415–422. doi: 10.1037/a0026884

Gross, S., and Rey, G. (2012). "Innateness" in The Oxford handbook of philosophy of cognitive science (Oxford, UK: Oxford University Press), 318–360.

Haggard, P. (2017). Sense of agency in the human brain. *Nat. Rev. Neurosci.* 18, 196–207. doi: 10.1038/nrn.2017.14

Hart, D., and Edelstein, W. (1992). Self-understanding development in cross-cultural perspective. The self: Deflinitional and methodological issues, 291–322.

Hart, D., Field, N. P., Garfinkle, J. R., and Singer, J. L. (1997). Representations of self and other: a semantic space model. *J. Pers.* 65, 77–105. doi: 10.1111/j.1467-6494.1997.tb00530.x

Harter, S. (2012). The construction of the self: Developmental and sociocultural foundations. New York, USA: Guilford Publications.

Haslam, C., Jetten, J., Haslam, S. A., Pugliese, C., and Tonks, J. (2011). 'I remember therefore I am, and I am therefore I remember': exploring the contributions of episodic and semantic self-knowledge to strength of identity. *Br. J. Psychol.* 102, 184–203. doi: 10.1348/000712610X508091

Hayward, C. R., and Watson, R. (2010). Identity and political theory Wash. *UJL Pol'y* 33:9.

Heeger, D. J. (2017). Theory of cortical function. *Proc. Natl. Acad. Sci.* 114, 1773–1782. doi: 10.1073/pnas.1619788114

Hepper, P. (2015). Behavior during the prenatal period: adaptive for development and survival. *Child Dev. Perspect.* 9, 38–43. doi: 10.1111/cdep.12104

Hess, E. H. (1959). Imprinting: an effect of early experience, imprinting determines later social behavior in animals. *Science* 130, 133–141. doi: 10.1126/science.130.3368.133

Heyes, C. (2018). Cognitive gadgets: The cultural evolution of thinking. Harvard, USA: Harvard University Press.

Hinsch, M., and Komdeur, J. (2017). What do territory owners defend against? *Proc. R. Soc. B* 284:20162356. doi: 10.1098/rspb.2016.2356

Hoffmann, M. (2022). Learning body models: from humans to humanoids. *arXiv Preprint*. arXiv:2211.03049. doi: 10.48550/arXiv.2211.03049

Hohwy, J. (2007). The sense of self in the phenomenology of agency and perception. *Psyche* 13, 1–20.

Hohwy, J. (2013). The predictive mind. *First edition*. Edn. Oxford, UK: Oxford University Press.

Hohwy, J., and Michael, J. (2017). Why should any body have a self? In VignemontF. de and A. Alsmith (Eds.), The Subject's matter. Self-Consciousness and the Body. MIT Press.

Hoppitt, W. J., Brown, G. R., Kendal, R., Rendell, L., Thornton, A., Webster, M. M., et al. (2008). Lessons from animal teaching. *Trends Ecol. Evol.* 23, 486–493. doi: 10.1016/j.tree.2008.05.008

Hu, C., Di, X., Eickhoff, S. B., Zhang, M., Peng, K., Guo, H., et al. (2016). Distinct and common aspects of physical and psychological self-representation in the brain: a meta-analysis of self-bias in facial and self-referential judgements. *Neurosci. Biobehav. Rev.* 61, 197–207. doi: 10.1016/j.neubiorev.2015.12.003

Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154. doi: 10.1113/jphysiol.1962.sp006837

Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195, 215–243. doi: 10.1113/jphysiol.1968.sp008455

Huddy, L. (2001). From social to political identity: a critical examination of social identity theory. *Polit. Psychol.* 22, 127–156. doi: 10.1111/0162-895X.00230

James, W. (1890). *The principles of psychology*. New York, USA: H. Holt and Company.

James, D., Pillai, M., and Smoleniec, J. (2013). Neurobehavioral development in the human fetus. In: *Fetal Development*. eds. J. Lecanuet, W. P. Fifer, N. A. Krasnegor and W. P. Smotherman (New York, USA: Psychology Press), 101–128.

Jenke, L., and Huettel, S. A. (2016). Issues or identity? Cognitive foundations of voter choice. *Trends Cogn. Sci.* 20, 794–804. doi: 10.1016/j.tics.2016.08.013

Johnson, M. H. (2005). Subcortical face processing. *Nat. Rev. Neurosci.* 6, 766–774. doi: 10.1038/nrn1766

Johnson, M. H., Dziurawiec, S., Ellis, H., and Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* 40, 1–19. doi: 10.1016/0010-0277(91)90045-6

Jones, M., and Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behav. Brain Sci.* 34, 169–188. doi: 10.1017/S0140525X10003134

Kalin, M., and Sambanis, N. (2018). How to think about social identity. *Annu. Rev. Polit. Sci.* 21, 239–257. doi: 10.1146/annurev-polisci-042016-024408

Kaplan, G. (2015). Bird minds: Cognition and behaviour of Australian native birds. Clayton, Australia: CSIRO Publishing.

Kaufman, J. C., and Sternberg, R. J. (2010). The Cambridge handbook of creativity. Cambridge, UK: Cambridge University Press.

Keller, G. B., and Mrsic-Flogel, T. D. (2018). Predictive processing: a canonical cortical computation. *Neuron* 100, 424–435. doi: 10.1016/j.neuron.2018.10.003

Kemp, C., and Tenenbaum, J. B. (2008). The discovery of structural form. *Proc. Natl. Acad. Sci. USA* 105, 10687–10692. doi: 10.1073/pnas.0802631105

Kenward, B. (2010). 10-month-olds visually anticipate an outcome contingent on their own action. *Infancy* 15, 337–361. doi: 10.1111/j.1532-7078.2009.00018.x

Kihlstrom, J. F., Beer, J. S., and Klein, S. B. (2003). “Self and identity as memory” in Handbook of self and identity. eds. M. R. T. Leary and J. Price (New York, USA: Guilford Press), 68–90.

Kim, K., and Johnson, M. K. (2014). Extended self: spontaneous activation of medial prefrontal cortex by objects that are 'mine'. *Soc. Cogn. Affect. Neurosci.* 9, 1006–1012. doi: 10.1093/scan/nst082

Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., et al. (2003). Effects of experience on fetal voice recognition. *Psychol. Sci.* 14, 220–224. doi: 10.1111/1467-9280.02435

Klein, S. B., German, T. P., Cosmides, L., and Gabriel, R. (2004). A theory of autobiographical memory: necessary components and disorders resulting from their loss. *Soc. Cogn.* 22, 460–490. doi: 10.1521/soco.22.5.460.50765

Knill, D. C., and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719. doi: 10.1016/j.tins.2004.10.007

Knoblich, G., Ohlsson, S., Haider, H., and Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *J. Exp. Psychol. Learn. Mem. Cogn.* 25:1534.

Kostović, I., and Judaš, M. (2010). The development of the subplate and thalamocortical connections in the human foetal brain. *Acta Paediatr.* 99, 1119–1127. doi: 10.1111/j.1651-2227.2010.01811.x

Kriegeskorte, N. (2015). Deep neural networks: a new framework for modeling biological vision and brain information processing. *Ann. Rev. Vision Sci.* 1, 417–446. doi: 10.1146/annurev-vision-082114-035447

Kumaran, D., Hassabis, D., and McClelland, J. L. (2016). What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends Cogn. Sci.* 20, 512–534. doi: 10.1016/j.tics.2016.05.004

Lake, B. M., Lawrence, N. D., and Tenenbaum, J. B. (2018). The emergence of organizing structure in conceptual representation. *Cogn. Sci.* 42, 809–832. doi: 10.1111/cogs.12580

Legrand, D., and Ruby, P. (2009). What is self-specific? Theoretical investigation and critical review of neuroimaging results. *Psychol. Rev.* 116, 252–282. doi: 10.1037/a0014172

Letheby, C., and Gerrans, P. (2017). Self unbound: ego dissolution in psychedelic experience. *Neurosci. Consciousness* 2017, 1–11. doi: 10.1093/nc/nix016

Levine, L. E. (1983). Mine: self-definition in 2-year-old boys. *Dev. Psychol.* 19, 544–549. doi: 10.1037/0012-1649.19.4.544

Lewis, M., and Brooks-Gunn, J. (1979). Toward a theory of social cognition: the development of self. *New Dir. Child Adolesc. Dev.* 1979, 1–20. doi: 10.1002/cd.23219790403

Lewis, P. A., Knoblich, G., and Poe, G. (2018). How memory replay in sleep boosts creative problem-solving. *Trends Cogn. Sci.* 22, 491–503. doi: 10.1016/j.tics.2018.03.009

Lewis, M., and Ramsay, D. (2004). Development of self-recognition, personal pronoun use, and pretend play during the 2nd year. *Child Dev.* 75, 1821–1831. doi: 10.1111/j.1467-8624.2004.00819.x

Limanowski, J. (2022). Precision control for a flexible body representation. *Neurosci. Biobehav. Rev.* 134:104401. doi: 10.1016/j.neubiorev.2021.10.023

Limanowski, J., and Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Front. Hum. Neurosci.* 7:547. doi: 10.3389/fnhum.2013.00547

Lorenz, K. (1935). Der Kumpan in der Umwelt des Vogels. Der Artgenosse als auslösendes Moment sozialer Verhaltungsweisen. *Journal für Ornithologie.* 83, 137–413.

Love, B. C., Medin, D. L., and Gureckis, T. M. (2004). SUSTAIN: a network model of category learning. *Psychol. Rev.* 111, 309–332. doi: 10.1037/0033-295X.111.2.309

Markus, H. (1977). Self-schemata and processing information about the self. *J. Pers. Soc. Psychol.* 35, 63–78. doi: 10.1037/0022-3514.35.2.63

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. San Francisco: WH Freeman.

Martin, J. R., and Pacherie, E. (2013). Out of nowhere: thought insertion, ownership and context-integration. *Conscious. Cogn.* 22, 111–122. doi: 10.1016/j.concog.2012.11.012

Martinelli, P., Sperduti, M., and Piolino, P. (2013). Neural substrates of the self-memory system: new insights from a meta-analysis. *Hum. Brain Mapp.* 34, 1515–1529. doi: 10.1002/hbm.22008

McAdams, D. P. (2001). The psychology of life stories. *Rev. Gen. Psychol.* 5, 100–122. doi: 10.1037/1089-2680.5.2.100

McAdams, D. P., and McLean, K. C. (2013). Narrative identity. *Curr. Dir. Psychol. Sci.* 22, 233–238. doi: 10.1177/0963721413475622

McCall, G. J., and Simmons, J. L. (1966). *Identities and interactions*. New York, USA: Free Press.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., et al. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends Cogn. Sci.* 14, 348–356. doi: 10.1016/j.tics.2010.06.002

McClelland, J. L., and Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nat. Rev. Neurosci.* 4, 310–322. doi: 10.1038/nrn1076

McCracken, G. F. (1984). Communal nursing in Mexican free-tailed bat maternity colonies. *Science* 223, 1090–1091. doi: 10.1126/science.223.4640.1090

McCracken, G. F., and Gustin, M. K. (1991). Nursing behavior in Mexican free-tailed bat maternity colonies. *Ethology* 89, 305–321. doi: 10.1111/j.1439-0310.1991.tb00376.x

Mead, G. H. (1934). Mind, self and society. Chicago, USA: University of Chicago Press.

Melzack, R., Israel, R., Lacroix, R., and Schultz, G. (1997). Phantom limbs in people with congenital limb deficiency or amputation in early childhood. *Brain J. Neurol.* 120, 1603–1620. doi: 10.1093/brain/120.9.1603

Mesoudi, A. (2011). Cultural evolution: How Darwinian theory can explain human culture and synthesize the social sciences. Chicago, USA: University of Chicago Press.

Miyazaki, M., Takahashi, H., Rolf, M., Okada, H., and Omori, T. (2014). The image-scratch paradigm: a new paradigm for evaluating infants' motivated gaze control. *Sci. Rep.* 4:5498. doi: 10.1038/srep05498

Moore, J. K., and Linthicum, F. H.Jr. (2007). The human auditory system: a timeline of development. *Int. J. Audiol.* 46, 460–478. doi: 10.1080/14992020701383019

Moutoussis, M., Fearon, P., El-Deredy, W., Dolan, R. J., and Friston, K. J. (2014). Bayesian inferences about the self (and others): a review. *Conscious. Cogn.* 25, 67–76. doi: 10.1016/j.concog.2014.01.009

Nancekivell, S. E., Friedman, O., and Gelman, S. A. (2018). Ownership matters: people possess a Naïve theory of ownership. *Trends Cogn. Sci.* 23, 102–113. doi: 10.1016/j.tics.2018.11.008

Navarro, D. J., and Kemp, C. (2017). None of the above: a Bayesian account of the detection of novel categories. *Psychol. Rev.* 124, 643–677. doi: 10.1037/rev0000077

Neisser, U. (1988). Five kinds of self-knowledge. *Philos. Psychol.* 1, 35–59. doi: 10.1080/09515088808572924

Nelson, K., and Fivush, R. (2004). The emergence of autobiographical memory: a social cultural developmental theory. *Psychol. Rev.* 111, 486–511. doi: 10.1037/0033-295X.111.2.486

Nelson, C. A.III, Zeanah, C. H., Fox, N. A., Marshall, P. J., Smyke, A. T., and Guthrie, D. (2007). Cognitive recovery in socially deprived young children: the Bucharest early intervention project. *Science* 318, 1937–1940. doi: 10.1126/science.1143921

Noble, G. K. (1939). The role of dominance in the social life of birds. *Auk* 56, 263–273. doi: 10.2307/4079047

Olshausen, B. A., and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381, 607–609. doi: 10.1038/381607a0

Orbán, G., Berkes, P., Fiser, J., and Lengyel, M. (2016). Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron* 92, 530–543. doi: 10.1016/j.neuron.2016.09.038

Owens, T. J., Robinson, D. T., and Smith-Lovin, L. (2010). Three faces of identity. *Ann. Rev. Sociol.* 36, 477–499. doi: 10.1146/annurev.soc.34.040507.134725

Park, H. D., and Tallon-Baudry, C. (2014). The neural subjective frame: from bodily signals to perceptual consciousness. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 369:20130208. doi: 10.1098/rstb.2013.0208

Peleg, D., and Goldman, J. A. (1980). Fetal heart rate acceleration in response to light stimulation äs a clinical measure of fetal well-being. A preliminary report. *J. Perinat. Med.* 8, 38–41

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., and Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition* 120, 302–321. doi: 10.1016/j.cognition.2010.11.015

Pezzulo, G., van der Meer, M. A., Lansink, C. S., and Pennartz, C. M. (2014). Internally generated sequences in learning and executing goal-directed behavior. *Trends Cogn. Sci.* 18, 647–657. doi: 10.1016/j.tics.2014.06.011

Piaget, J. (2013). The construction of reality in the child. England, UK: Routledge.

Plumptre, A. J., and Reynolds, V. (1997). Nesting behavior of chimpanzees: implications for censuses. *Int. J. Primatol.* 18, 475–485. doi: 10.1023/A:1026302920674

Pothos, E. M., and Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cogn. Sci.* 26, 303–343. doi: 10.1207/s15516709cog2603_6

Povinelli, D. J., Landau, K. R., and Perilloux, H. K. (1996). Self-recognition in young children using delayed versus live feedback: evidence of a developmental asynchrony. *Child Dev.* 67, 1540–1554. doi: 10.2307/1131717

Ralph, M. A., Jefferies, E., Patterson, K., and Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* 18, 42–55. doi: 10.1038/nrn.2016.150

Ramachandran, V., Altschuler, E., and Hillyer, S. (1997). Mirror agnosia. *Proc. R. Soc. Lond. B Biol. Sci.* 264, 645–647. doi: 10.1098/rspb.1997.0091

Ramsey, W. (2016). Untangling two questions about mental representation. *New Ideas Psychol.* 40, 3–12. doi: 10.1016/j.newideapsych.2015.01.004

Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Reid, V. M., Dunn, K., Young, R. J., Amu, J., Donovan, T., and Reissland, N. (2017). The human fetus preferentially engages with face-like visual stimuli. *Curr. Biol.* 27:2052. doi: 10.1016/j.cub.2017.06.036

Riva, G. (2018). The neuroscience of body memory: from the self through the space to the others. *Cortex* 104, 241–260. doi: 10.1016/j.cortex.2017.07.013

Robertson, S. S. (1990). "Temporal Organizaton in fetal and newborn movement" in *Sensory-Motor Organizations and Development in Infancy and Early Childhood: Proceedings of the NATO Advanced Research Workshop on Sensory-Motor Organizations and Development in Infancy and Early Childhood Chateu de Rosey, France* (Netherlands: Springer), 105–122.

Rochat, P. (2011). Possession and morality in early development. *New Dir. Child Adolesc. Dev.* 2011, 23–38. doi: 10.1002/cd.294

Rochat, P., and Striano, T. (1999). Emerging self-exploration by 2-month-old infants. *Dev. Sci.* 2, 206–218. doi: 10.1111/1467-7687.00069

Rochat, P., and Striano, T. (2000). Perceived self in infancy. *Infant Behav. Dev.* 23, 513–530. doi: 10.1016/S0163-6383(01)00055-8

Rochat, P., and Zahavi, D. (2011). The uncanny mirror: a re-framing of mirror self-experience. *Conscious. Cogn.* 20, 204–213. doi: 10.1016/j.concog.2010.06.007

Rodgon, M. M., and Rashman, S. E. (1976). Expression of owner-owned relationships among holophrastic 14-to 32-month-old children. *Child Dev.* 47, 1219–1222. doi: 10.2307/1128470

Salomon, R. (2017). The assembly of the self from sensory and motor foundations. *Soc. Cogn.* 35, 87–106. doi: 10.1521/soco.2017.35.2.87

Samuels, R. (2012). "Massive modularity" in *The Oxford handbook of philosophy of cognitive science*. eds. E. Margolis, R. Samuels and S. P. Stich, (Oxford, UK: Oxford University Press).

Schechtman, M. (2011). "The narrative self" in The Oxford handbook of the self. ed. S. Gallagher (Oxford, UK: Oxford University Press).

Schillaci, G., Ritter, C.-N., Hafner, V. V., and Lara, B. (2016). Body representations for robot ego-noise modelling and prediction. Towards the development of a sense of agency in artificial agents. *ALIFE 2016, the Fifteenth International Conference on the Synthesis and Simulation of Living Systems*. Cambridge, USA: MIT Press. 390–397.

Schulman, A. H., and Kaplowitz, C. (1977). Mirror-image response during the first two years of life. *Develop. Psychobiol.* 10, 133–142. doi: 10.1002/dev.420100206

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends Cogn. Sci.* 17, 565–573. doi: 10.1016/j.tics.2013.09.007

Seth, A. K., Suzuki, K., and Critchley, H. D. (2011). An interoceptive predictive coding model of conscious presence. *Front. Psychol.* 2:395. doi: 10.3389/fpsyg.2011.00395

Seth, A. K., and Tsakiris, M. (2018). Being a beast machine: the somatic basis of selfhood. *Trends Cogn. Sci.* 22, 969–981. doi: 10.1016/j.tics.2018.08.008

Shipp, S. (2016). Neural elements for predictive coding. *Front. Psychol.* 7:1792. doi: 10.3389/fpsyg.2016.01792

Simoncelli, E. P., and Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annu. Rev. Neurosci.* 24, 1193–1216. doi: 10.1146/annurev.neuro.24.1.1193

Sims, A. C. (2017). The problems with prediction: the dark room problem and the scope dispute. In: *Philosophy and Predictive Processing* eds. T. Metzinger and W. Wiese. (Frankfurt am Main, Germany: Mind Group), pp. 1–18.

Smith, E. R., Coats, S., and Walling, D. (1999). Overlapping mental representations of self, in-group, and partner: further response time evidence and a connectionist model. *Personal. Soc. Psychol. Bull.* 25, 873–882. doi: 10.1177/0146167299025007009

Smith, R., Schwartenbeck, P., Parr, T., and Friston, K. J. (2019). An active inference model of concept learning. *bioRxiv*:633677. doi: 10.1101/633677

Spratling, M. W. (2010). Predictive coding as a model of response properties in cortical area V1. *J. Neurosci.* 30, 3531–3543. doi: 10.1523/JNEUROSCI.4911-09.2010

Spratling, M. W. (2016). Predictive coding as a model of cognition. *Cogn. Process.* 17, 279–305. doi: 10.1007/s10339-016-0765-6

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain Cogn.* 112, 92–97. doi: 10.1016/j.bandc.2015.11.003

Stamps, J. (1994). Territorial behavior: testing the assumptions. *Adv. Study Behav.* 23:232.

Sternberg, R. J., and Davidson, J. E. (1995). The nature of insight. Cambridge: The MIT Press.

Stipek, D. J., Gralinski, J. H., and Kopp, C. B. (1990). Self-concept development in the toddler years. *Dev. Psychol.* 26, 972–977. doi: 10.1037/0012-1649.26.6.972

Strassmann, J. E., and Queller, D. C. (2014). Privatization and property in biology. *Anim. Behav.* 92, 305–311. doi: 10.1016/j.anbehav.2014.02.011

Stryker, S. (1968). Identity salience and role performance: the relevance of symbolic interaction theory for family research. *J. Marriage Fam.* 30, 558–564. doi: 10.2307/349494

Stryker, S. (2008). From Mead to a structural symbolic interactionism and beyond. *Ann. Rev. Sociol.* 34, 15–31. doi: 10.1146/annurev.soc.34.040507.134649

Suddendorf, T., and Butler, D. L. (2013). The nature of visual self-recognition. *Trends Cogn. Sci.* 17, 121–127. doi: 10.1016/j.tics.2013.01.004

Suddendorf, T., Simcock, G., and Nielsen, M. (2007). Visual self-recognition in mirrors and live videos: evidence for a developmental asynchrony. *Cogn. Dev.* 22, 185–196. doi: 10.1016/j.cogdev.2006.09.003

Sui, J., and Gu, X. (2017). Self as object: emerging trends in self research. *Trends Neurosci.* 40, 643–653. doi: 10.1016/j.tins.2017.09.002

Sui, J., and Humphreys, G. W. (2015). The integrative self: how self-reference integrates perception and memory. *Trends Cogn. Sci.* 19, 719–728. doi: 10.1016/j.tics.2015.08.015

Tajfel, H. (1982). Social-psychology of inter-group relations. *Annu. Rev. Psychol.* 33, 1–39. doi: 10.1146/annurev.ps.33.020182.000245

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science* 331, 1279–1285. doi: 10.1126/science.1192788

Thornton, A., and Raihani, N. J. (2010). Identifying teaching in wild animals. *Learn. Behav.* 38, 297–309. doi: 10.3758/LB.38.3.297

Tobita, M., Hasegawa, O., Nagatomo, H., Yamaguchi, S., and Kurita, R. (1995). Autotopagnosia ameliorated by looking at the image reflected in a mirror. Rinsho shinkeigaku=. *Clin. Neurol.* 35, 296–298

Tomasello, M. (1998). One child's early talk about possession. *Typol. Stud. Lang.* 36, 349–373. doi: 10.1075/tsl.36.13tom

Tomasello, M., Kruger, A. C., and Ratner, H. H. (1993). Cultural learning. *Behav. Brain Sci.* 16, 495–511. doi: 10.1017/S0140525X0003123X

Truong, G., and Todd, R. M. (2017). SOAP opera: self as object and agent in prioritizing attention. *J. Cogn. Neurosci.* 29, 937–952. doi: 10.1162/jocn_a_01083

Ullman, T. D., and Tenenbaum, J. B. (2020). Bayesian models of conceptual development: learning as building models of the world. *Ann. Rev. Dev. Psychol.* 2, 533–558. doi: 10.1146/annurev-devpsych-121318-084833

Van Bavel, J. J., and Pereira, A. (2018). The partisan brain: an identity-based model of political belief. *Trends Cogn. Sci.* 22, 213–224. doi: 10.1016/j.tics.2018.01.004

Verret, J., and Lapresle, J. (1978). Anton-Babinski syndrome with recognition of the left upper limb on visualization in a mirror. *Rev. Neurol.* 134, 709–713.

Villarejo, A., Martin, V. P., Moreno-Ramos, T., Camacho-Salas, A., Porta-Etessam, J., and Bermejo-Pareja, F. (2011). Mirrored-self misidentification in a patient without dementia: evidence for right hemispheric and bifrontal damage. *Neurocase* 17, 276–284. doi: 10.1080/13554794.2010.498427

Wang, Q. (2004). The emergence of cultural self-constructs: autobiographical memory and self-description in European American and Chinese children. *Dev. Psychol.* 40, 3–15. doi: 10.1037/0012-1649.40.1.3

Watanabe, H., and Taga, G. (2006). General to specific development of movement patterns and memory for contingency between actions and events in young infants. *Infant Behav. Dev.* 29, 402–422. doi: 10.1016/j.infbeh.2006.02.001

Watanabe, H., and Taga, G. (2009). Flexibility in infant actions during arm-and leg-based learning in a mobile paradigm. *Infant Behav. Dev.* 32, 79–90. doi: 10.1016/j.infbeh.2008.10.003

Watanabe, H., and Taga, G. (2011). Initial-state dependency of learning in young infants. *Hum. Mov. Sci.* 30, 125–142. doi: 10.1016/j.humov.2010.07.003

Waters, E., Matas, L., and Sroufe, L. A. (1975). Infants' reactions to an approaching stranger: description, validation, and functional significance of wariness. *Child Dev.* 46, 348–356. doi: 10.2307/1128127

Watson, J. S., and Ramey, C. T. (1972). Reactions to response-contingent stimulation in early infancy. *Merrill-Palmer Quarterly Behav. Dev.* 18, 219–227.

Weisberg, R. W. (2015). Toward an integrated theory of insight in problem solving. *Think. Reason.* 21, 5–39. doi: 10.1080/13546783.2014.886625

Whitehouse, H. (2018). Dying for the group: towards a general theory of extreme self-sacrifice. *Behav. Brain Sci.* 41:e192. doi: 10.1017/S0140525X18000249

Williams, D. (2018). Predictive processing and the representation wars. *Mind. Mach.* 28, 141–172. doi: 10.1007/s11023-017-9441-6

Woźniak, M. (2018). "I" and "me": the self in the context of consciousness. *Front. Psychol.* 9:1656. doi: 10.3389/fpsyg.2018.01656

Yamins, D. L., and DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* 19, 356–365. doi: 10.1038/nn.4244

Young, G. (2008). On how a child's awareness of thinking informs explanations of thought insertion. *Conscious. Cogn.* 17, 848–862. doi: 10.1016/j.concog.2007.05.005

Yuill, N. (1992a). Children's production and comprehension of trait terms. *Br. J. Dev. Psychol.* 10, 131–142. doi: 10.1111/j.2044-835X.1992.tb00567.x

Yuill, N. (1992b). Children's conception of personality traits. *Hum. Dev.* 35, 265–279. doi: 10.1159/000277220

Zaadnoordijk, L., Meyer, M., Zaharieva, M., Kemalasari, F., van Pelt, S., and Hunnius, S. (2020). From movement to action: an EEG study into the emerging sense of agency in early infancy. *Dev. Cogn. Neurosci.* 42:100760. doi: 10.1016/j.dcn.2020.100760