Body Schema in Robotics: A Review

Matej Hoffmann, Hugo Gravato Marques, Alejandro Hernandez Arieta, Hidenobu Sumioka, Max Lungarella, and Rolf Pfeifer

Abstract—How is our body imprinted in our brain? This seemingly simple question is a subject of investigations of diverse disciplines, psychology, and philosophy originally complemented by neurosciences more recently. Despite substantial efforts, the mysteries of body representations are far from uncovered. The most widely used notions—body image and body schema—are still waiting to be clearly defined. The mechanisms that underlie body representations are coresponsible for the admiring capabilities that humans or many mammals can display: combining information from multiple sensory modalities, controlling their complex bodies, adapting to growth, failures, or using tools. These features are also desirable in robots. This paper surveys the body representations in biology from a functional or computational perspective to set ground for a review of the concept of body schema in robotics. First, we examine application-oriented research: how a robot can improve its capabilities by being able to automatically synthesize, extend, or adapt a model of its body. Second, we summarize the research area in which robots are used as tools to verify hypotheses on the mechanisms underlying biological body representations. We identify trends in these research areas and propose future research directions.

Index Terms—Body image, body representation, body schema, forward model, robotics, self-calibration.

I. INTRODUCTION

THE basic notion of body schema encloses a group of body representations which are essential for body motion and a meaningful interaction with the environment carried out by an embodied agent. The body schema allows for integration of information from proprioception, vision, audition, the vestibular system, tactile sensing, and from the motor system in order to keep an up-to-date representation of the positions of the different body parts in space. Typically, these representations are involved in movement preparation and the representation of space in different frames of reference to be used by different behaviors. Such representations are the central subject of many studies in the cognitive sciences, especially in the neurosciences. The concepts of a postural schema and a surface schema were first introduced by Head and Holmes [1]. In their view, the postural schema represents the awareness we have of our bodies' position in space, and the surface schema represents our capacity to locate stimuli on the surface of the skin. Since then, many other classifications and taxonomies appeared trying to structure the plethora of body representations; yet up to now, the literature has not converged to any of them.

Biological agents are able to adapt seamlessly to new situations or cope with failures. To a large extent, this is because the body representations required to support their behaviors can dynamically adapt to new circumstances. These properties are also desirable in robots; today, their operation is still restricted to static or limited environments, and resilience to failure is typically absent. When trying to bridge this gap, many roboticists look to biology for inspiration to integrate some of the features of a biological body schema into their machines. While this flow of information—from biology to robotics—was dominant so far, there also exists a route in the opposite direction. Many of the mechanisms underlying the body schema are still a mystery to cognitive scientists. Here, robots can qualify as useful tools to test hypotheses that have been put forward up to this day. In particular, although there is rapid progress, to a large extent thanks to the neurosciences with their imaging techniques, the investigation of some mechanisms requires whole brain–body–environment systems as test-beds. Experiments on robots can thus complement the research in computational neuroscience.

This paper is structured as follows. First, we offer a review of body representations in the context of biology. After discussing taxonomies of body representations and how these are supported by studies on disorders, we will focus on topics that we consider of greatest relevance for robotics: plasticity of body representations (development, adaptation, extension), coordinate transformations, and the relationship between body schema and forward models. Second, we provide an overview of the engineering-oriented work in which a body schema serves to control a robot and to improve its behavior when faced with unexpected circumstances. In theory, an enormous part of research in robotics and control could fall into this section, since models (plant models) used to control robots are ubiquitous. However, we will show how this representation is different from the ones that take inspiration from biology and further concentrate on the latter. There are many axes according to which the research in robotics could be structured. For us, the principle axis will be the nature of representation: explicit versus implicit. Third, a section on robots employed as tools to model biological body representations is presented. We think that this body of work—investigating whole brain-body-environment systems—is a necessary complement of computational neuroscience. We conclude by identifying the major trends and suggesting future research directions.

II. BODY REPRESENTATIONS IN BIOLOGY

Significant evidence has been accumulated up to this day testifying that there are representations of the body in the brain. It is also very likely that there is no single unitary representation, but rather several, partial representations that serve different purposes. We will discuss the basic taxonomies of body representations and define the two most widely used notions: body schema and body image. Disorders and dissociation studies are useful to get insight into the structure of the putative body representations. We will discuss two in detail and present an overview of those that we consider relevant for robotics.

Body representations are plastic over time. This property is largely responsible for many of the capabilities that animals display. We will discuss the developmental time scale first. How does an infant acquire its body representation? How does it develop a sense of body ownership and agency? Then, we will review the plasticity of body representations over short time scales, minutes for instance. We will examine the "rubber hand illusion" and extension of body representations in tool use. The next topic of undisputed direct relevance for robotics are coordinate transformations. Finally, we have included a section which demonstrates the notions discussed on a concrete scenario. We also establish an explicit relation to forward internal models—a closely related concept. The idea is to provide enough information for a roboticist to get an initial functional understanding of the topic and to equip her with initial pointers to the literature. We have to admit that this section is strongly biased toward body representations in humans and primates. Other animals remained out of the scope of this review. However, studying body representations in simpler animals than humans can provide no less valuable insights for roboticists.

A. What Is a Body Schema?

Two main taxonomies form a first attempt to differentiate the variety of body representations: the dyadic and the triadic taxonomies [2]. Both draw a line between representations that are used for action and those used for perception. This functional division is grounded on the hypothesis that visual as well as somatosensory processing is carried out in two distinct nervous pathways: one for action and another for conscious perception and object recognition [3]–[5]. In visual processing, these are the "what" and "how" streams as suggested in [6] (earlier distinction between "what" and "where" pathways was suggested in [7]).

The visual pathway for action, the "how" or dorsal stream, goes from the occipital lobe to the motor cortex through the parietal cortex. The pathway for perception, the "what" or ventral stream, goes from the occipital lobe to the temporal lobe. A similar separation can be observed in somatosensory perception. The pathway for action involves the anterior parietal cortex (APC), eventually the secondary somatosensory area (SII) and terminates in the posterior parietal cortex (PPC) [4]; the pathway for perception involves a similar route but terminates at the insula rather than at the PPC. The right PPC might also be involved when integration of spatio–temporal information is required for the recognition of objects as well as body configuration (see also [8] and [4], and [6] for a comparison between the two pathways for action and perception in the somatosensory processing and visual and auditory processing).

On these grounds, the *dyadic taxonomy* distinguishes between *body schema* and *body image*. The former are sensorimotor representations of the body used to guide movement and action, the latter are used to form our perceptual (body percept), conceptual (body concept) or emotional (body affect) judgments towards our body [3]. However, especially the concept of body image is problematic, lacking a positive definition; it seems that once we are done with a body schema, everything else can fall into body image [2]. Therefore, the *triadic taxonomy* further splits the representations belonging to the general concept of body image [9]. One of these representations, the body structural description, entails a topological representation (mainly visual) of the position of the different body parts in relation to each other (e.g., the forearm extends the upper arm via a hinge joint). The other representation, body semantics, comprises a semantic representation of the body which includes the names of the different body parts, their functions as well as potential relations to external artifacts (e.g., shoes are used on the feet, and feet can be used to kick a football).

The functional axis, action versus perception, is only one possible criterion to distinguish between various body representations, and an oversimplifying one for that matter. Other features used to classify body representations are availability to consciousness (unconscious versus conscious), and dynamics (short-term versus long-term). However, the weight of the criteria varies relative to the author and sometimes even the same notion is ascribed opposite properties (see [2] for more details). While these additional axes are useful, they still do not provide any clear taxonomy of body representations. Perhaps, such an endeavor cannot be successful, because we are not faced with two or three distinct representations, but rather with a panoply of many interacting partial representations.

Nevertheless, there is definitely some agreement that there is something like a body schema—a sensorimotor representation of the body used for action. Typically, it would not be available to consciousness[1] and would encompass both short-term (e.g., position of a limb at a given instant) and long-term dynamics (e.g., biomechanical properties and size of limbs). Since we are mainly interested in body representations in robots, this notion will be our primary focus. Our decision is motivated by the following reasons: 1) as stated above, there is certain consensus on the existence of a body schema; 2) the fact that it is a representation for action finds a natural counterpart in robots which can then be employed to perform tasks; and 3) we think that robots have not yet reached the level of competence where notions like conscious representations can be investigated in a grounded fashion.

The notion of body image—as a perception-based representation—will not be excluded from our investigation; however, we will restrict it to the body structural or topological representation, leaving apart the domains of body concept or body affect.

---

[1]Though it may become conscious under certain circumstances, such as during motor imagery [9].

B. Disorders

What are the grounds on which the body representations are classified into the taxonomies we have come across? Underpinning these taxonomies are a variety of studies which analyze the functional impact of some impairment on the behavior of a subject (see [4]). It is the fact that some subjects are able to perform normally on some body-related tasks but not on others that allows to distinguish between the different representations.

Probably the most mentioned disorder in the context of research on body schema is that of deafferentation. Deafferentation in general is the (total or partial) deterioration of afferent signals, i.e., signals that go from the periphery to the central nervous system. When applied to body-related representations, deafferentation is the (complete or partial) loss of proprioceptive and tactile signals whether their origin is in the periphery or in more central areas. Paillard [5] reported two cases of deafferented patients with very different behaviors. In one case, the patient G.L. was able to perceive a signal applied to her body, report verbally the location of the stimulus as well as to point to the correct location of the limb part stimulated on a body sketch. However, when asked to point with her right hand to the part of her own body which had been stimulated, she was unable to do so. In the other (somehow more bizarre) case, patient R.S. was unable to consciously perceive tactile stimuli, joint positioning, temperature or pain in her own body; she could for example cut or burn herself without noticing. R.S. failed to locate verbally a given tactile stimuli on her own body but curiously (even to herself) she could point flawlessly to the body part stimulated. According to Paillard, these two cases provide a case for an intact body schema with an impaired body image (R.S.) and a case for an impaired body schema with an intact body image (G.L.); i.e., they provide a case for the distinction between the two body representations in the brain as mentioned in the previous section [3].

Cases showing a further distinction between body structural description and body semantics can also be found in the literature. In a large group study, Schwoebel and Coslett [9] analyzed subjects on three types of measures: one assessing the integrity of the body schema, one assessing the integrity of the body structural description, and another assessing the integrity of the body semantic representation. Each performance measure involved a set of different tasks [9]. In the first measure, aimed at assessing the integrity of the body schema, subjects were required either to: 1) imagine or execute different finger movements; or 2) to indicate the laterality of a hand in a picture (i.e., left or right hand). The second measure, aimed at assessing the body structural description, included three tasks: 1) to point to the location in one's own body of a body part depicted in an image; 2) to point to the location of a stimulus applied to a given body part; and 3) to point to one of three pictured body parts that were closer to a given target body surface. The third measure, aimed to assess the integrity of the body semantic description, involved two tasks: 1) to match one of three pictured body parts with another functionally related target part (e.g., the elbow has as similar function as the knee; they are both hinge joints); and 2) to match a pictured item of cloth with one of four given pictured body parts. They found out that 13 of the patients analyzed failed on tasks involving the measure of body schema integrity but performed normally on the other two measures. Three of the

TABLE I
DISORDERS RELATED TO THE BODY SCHEMA (EXTRACTED FROM [2])

Alice in Wonderland Syndrome	Distorted awareness of body size, mass, or its position in space
Allochiria	Mislocation of sensory stimuli to opposite half of the body
Anarchic hand sign	Unintended but purposeful movements of the upper limb and intermanual conflict
Autoscopy	Experience of seeing one's body in extrapersonal space
Autoprosopagnosia	Inability to recognize one's own face
Autotopagnosia	Mislocalisation of body parts and bodily sensations
Body form agnosia	Deficit of recognition of body parts
Body-specific aphasia	Loss of lexical knowledge of body parts
Deafferentation	Loss of tactile and proprioceptive information
Dysmorphophosia	Distorted perception of one's self-appearance
Fading limb	Lack of awareness of the presence and position of the limb if not seen
Finger agnosia	Inability to individuate and recognize the fingers
Gertmann's syndrome	Finger agnosia, agraphia, acalculia and left-right confusion
Heterotopagnosia	Designation of parts of the body of another person when asked to point towards one's own body
Ideomotor apraxia	Inability to execute or carry out skilled movements and gestures
Macro/microsomatognosia	Distorted awareness of body or body parts' size
Mirror sign	Inability to recognize one's own image in the mirror
Motion sickness (or kinetosis)	Vestibular balance disorder
Motor neglect	Underutilisation of one side of the body
Numbsense	Tactile deficit with preserved tactually guided movements
Out of body experience (OBE)	Visual awareness of one's own body from a location outside the physical body
Personal neglect	Lack of attention towards one's side of the body
Phantom limb	Awareness of an amputated limb
Pusher syndrome	Postural deviation towards the contralesional side
Prosopagnosia	Deficit of face recognition
Supernumerary limb	Awareness of non-existing limbs
Tactile extinction	Lack of awareness of tactile stimuli on the contralesional limb during simultaneous bilateral stimulation

patients failed to carry out successfully the tasks involved in the body structural description measure, but were able to carry out normally the tasks involved in the other two measures. Finally, two of the patients failed to execute the tasks related to the body semantic measures but performed normally in the other tasks. These results provide grounds to support the triadic taxonomy.

If robots are to be used as models of biological (in this case human) body representations, they can eventually be also subject to such tests—failures in robots can be compared to disorders in humans. A list with main disorders related to body representations is given in Table I. This list is a short version of the one offered in Vignemont [2]. The original table was pruned in order to give only the information most relevant for roboticists; the disorders removed were basically related to eating disorders or emotional responses related to body representations.

C. Plasticity of Body Representations

1) Development, Body Ownership, and Agency: How do the various body representations originate? They arise during the process of development immediately after birth (or even before—in the womb). We have to rely more on psychological rather than neurophysiological data here, since brain imaging techniques are not readily applicable on infants. As reported by Rochat [10], infants spend substantial time in their early months observing and touching themselves. Rochat calls it the visual–proprioceptive calibration of the body. Through this

process of babbling, intermodal redundancies, temporal contingencies, and spatial congruences are picked up. Environmental stimulation (single touch) can be distinguished from self-stimulation (double-touch + proprioceptive stimulation) [11]. If we treat this process as relying mainly on perception, we can view it as the acquisition of the body image. However, the infants not only observe, but actively involve their motor apparatus in the explorations (e.g., [12]). Hence, the development of body schema probably takes place at the same time.

Hand in hand with the development of the body representations, the infants acquire a notion of *body ownership* and *agency*. By sense of body ownership we mean that the infant knows that it is its body that is moving, even passively; sense of agency corresponds to the notion that the infant (or agent) knows that it is causing or generating an action. We mean agency in a low-level sense here—prereflective, sensorimotor, and functional, rather than in a phenomenological sense (see e.g., [13] for a disambiguation). Basically, a sense of body ownership would be disrupted by a sensory experience that does not match the previously learned regularities between modalities (i.e., mismatch with body image); sense of agency would be disrupted by a sensory–motor mismatch (i.e., a mismatch with a body schema).[2] However, as it is hard to separate body image from body schema, it is also hard to separate sense of body ownership from sense of agency (see [14] for details and experimental treatment of this issue). The above-mentioned low-level capabilities constitute the basis for action recognition in self, action recognition in others, and self-other discrimination. This is further related to action mirroring (where the mirror neurons are active) and imitation (see Rizzolatti *et al.* [15]). Such capabilities constitute a natural extension of our topic, but will remain largely out of the scope of this review.

2) Rubber Hand Illusion: The body representations are not only plastic during development. They can also respond to large changes in the body, such as limb loss (see e.g., Ramachandran and Blakeslee [16]). Moreover, body representations can also adapt over much shorter time scales. Let us first look how this can happen on the perceptual side—modifying the body image. Holmes and Spence [17] made an extensive review on different behavioral, neurophysiological, and neuropsychological studies regarding evidence on the possibility to "incorporate" objects not connected directly to the body by multisensory integration. A prominent series of studies, started by Botvinick and Cohen [18], involves the "rubber hand illusion." A subject looks at a rubber replica of her hand while her own hand is hidden. Through simultaneous tactile stimulation of the subject's hand and the rubber hand, visible on the rubber hand only, the rubber hand becomes incorporated into the body image and the subject is deceived to think she "owns" the rubber hand. In other words, simultaneous tactile stimulation, together with congruent visual and proprioceptive feedback, causes a rapid adaptation of the body image and the rubber hand enters our sense of body ownership. Graziano [19] reported a similar phenomenon in mon-

[2]In more concrete terms, this means there will be a mismatch between the sensory feedback predicted by a forward model from a motor command (efference) copy and the actual sensory input (reafference). This is also referred to as the "comparator model." However, for both body ownership and agency, the situation is more complicated and involves a top–down component as well—knowledge about the context [13], [14].
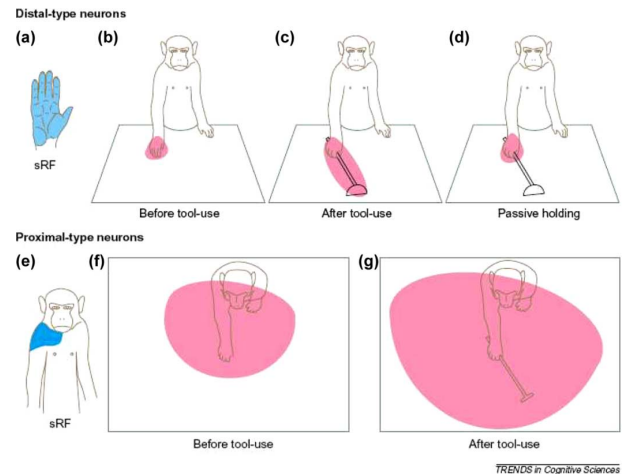

Fig. 1. Changes in bimodal receptive field properties following tool-use. The somatosensory receptive fields (sRF) of cells in this region were identified by light touches, passive manipulation of joints, or active hand-use. The visual RF (vRF) was defined as the area in which cellular responses were evoked by visual probes (the most effective ones being those moving towards the sRF). (a) sRF (blue area) of the "distal type" bimodal neurons and their vRF (pink areas) (b) before tool-use, (c) immediately after tool-use, and (d) when just passively grabbing the rake. (e) sRF (blue area) of "proximal type" bimodal neurons, and their vRF (pink areas), (f) before, and (g) immediately after tool-use. Reprinted from [31] with permission.

keys as well. More recently, other studies ([20]–[23]) further explored the rubber hand paradigm, also in the case of hand amputees [24]. The appropriation of the external object as part of the body representation of the person goes to the extent that if the rubber hand is threatened, the person shows a similar level of activity in the brain areas associated with anxiety and interoceptive awareness [25]. This effect can be found in the appropriation of virtual bodies as well [26].

3) Tool Use: Tsakiris *et al.* [14] pointed out that the basic rubber hand illusion setup lacks ecological validity, because it does not involve bodily movement. In other words, it is not a usual situation for primates or humans not to actively perform actions, but rely on multisensory integration only. Efferent information may play a key role, bringing us back to the difficulty of separation between body schema and image, or body ownership and agency. This leads us to another prominent experimental paradigm: *body schema extension* during *tool use* (body schema because now we are concerned with representations for action). Primates can manipulate objects in different ways and some can use tools to achieve a particular goal. Maravita and Iriki [27] investigated the integration of a tool into one's body schema in a macaque monkey that was retrieving food with the help of a rake. Neuronal activity of bimodal neurons (i.e., neurons that react to both somatosensory and visual stimulation) was recorded from the intraparietal cortex. Two groups of neurons were identified: "distal type" and "proximal type" (see Fig. 1). The former responded to somatosensory stimuli at the hand and visual stimuli near the hand. The visual receptive field (region of space in which the presence of a stimulus will alter the firing of a particular neuron) of these neurons followed the hand in space. After the monkey had used the tool for about five minutes, the visual receptive field of some neurons expanded to cover the entire length of the tool [Fig. 1(c)].

The visual receptive field of the latter neuron group—the "proximal type"—was not centered around the hand, but spanned the whole space within reach [Fig. 1(g)]. This space is called *peripersonal space*. As the body and the space immediately surrounding it are always in close interaction, the same seems to hold for their representations. Therefore, the representation of the body and of peripersonal space—space within reach—have to go hand in hand [28]. In the monkey, this space was expanded accordingly after working with the tool to accommodate the whole space that can be accessed with the tool.

Several studies [29]–[34] followed that show the ability of the primate brain to incorporate tools into its body representations and use them for coordinated action. The visual receptive field was extended by the tool when it was used for retrieving food, but not when the monkey just held the rake passively in its hand [Fig. 1(d)]. This confirms the hypothesis that action context plays a key role. It is also probable—unlike in the rubber hand illusion scenario—that the subject is not fully "deceived" to think that the tool is part of her body—the tool does not look like the hand—but only incorporates it into the representation in order to be able to use it as a "body auxiliary" [14].

4) Intelligent Tools: A tool can be much more than a passive rake. It can be an artifact with "intelligence" of its own. Slater *et al.* [35] show how it is possible to induce the incorporation of virtual bodies in the body representation. The advances in brain–computer interfaces [36]–[38] have made it possible to use biological signals to control robotic devices, enabling their users to perform activities otherwise out of their reach. These interfaces allow direct interaction with cortical processes that the user can control. So far evidence with monkeys [39], [40] show that they can "incorporate" intelligent devices into their body representation. Other studies with amputees [41]–[43] present evidence of changes in the cortical activation due to the interaction with an intelligent prosthetic hand.

Taking the "intelligence" of the artifact or device one step further, Sanchez *et al.* [44] and DiGiovanna [45] explore symbiotic systems where not only is the artifact incorporated in the body representation of its user, this time a rat, but at the same time, the intelligent artifact actively participates in the process. The artificial system taps into the user's brain and uses reinforcement learning to modify its own parameters in order to maximize the match between the user's intention and the action performed with the artifact. Thus, both the user and the tool coadapt to accomplish the task.

It is not hard to imagine how the plasticity of body representations, which was discussed in this section, can be useful for robots. A robot that can automatically acquire a model of itself that can then be used for control will save a lot of work to programmers. If it is able to automatically adapt the model to new circumstances—body extensions, wear and tear, or even substantial failures—it will lead to a new generation of robots which can leave their restricted work conditions.

D. Coordinate Transformations

A key issue that is often mentioned in the context of the body schema is the one of coordinate transformations. The problem is simple to formulate, but hard to tackle. Imagine you see an orange at some location in space and you want to grasp it. It might seem trivial for you to simply stretch your arm and reach it, but how is the brain successful at it? The orange falls on some location in the retina, which is dependent on the position of the eyes, the head, and the torso; if you move either of them (or all, as far as their movements do not cancel out) then the location of the orange in the retina will change accordingly. To perform a particular movement the brain has to have (in principle) at least one stable frame of reference (FoR), i.e., a FoR which is invariant to changes in the position of some of the body parts (say, the eyes or the head). A stable FoR for reaching is the torso frame of reference, since all the movements of the hand have to be necessarily executed in relation to the torso—due to the physical structure which connects the two body parts. However, to have the position of the orange encoded in relation to the torso, the brain has to convert first the retina coordinates into eye coordinates using the location of the orange on the retina, then transform the position in eye FoR to the head FoR using the current orientation of the eyes with respect to the head, and finally transform the location of the object with respect to the head into torso FoR using the orientation of the head in relation to the torso.

The brain areas which are often mentioned in the context of body schema and coordinate transformations are: the lateral intraparietal area (LIP), which encodes information relevant for saccadic eye movements [46], the ventral intraparietal area (VIP) which encodes both visual and somatosensory information [47], [48] and is connected to LIP area [49] and premotor areas responsible for head movements [50], the parietal reach region (PRR) which encodes reaching information [51], and the anterior parietal area which encodes grasping information. Each of these areas seems to use different frames of reference. This would be expected as different behaviors might benefit from a different encoding. For example, parts of LIP and VIP are supposed to represent the position of a visual target in both eye-centered and head-centered coordinate systems [47], [52]; neurons in the PRR should have the eye as their reference frame [51]. Interestingly, other neurons in the PRR have also been found which seem to encode the difference between a target in eye FoR and the current position of the hand also in eye FoR [53]. Such neurons seem to be particularly suited to output an error signal with the distance between the hand and the target [53]. Similar neurons have also been found in area 5 of the posterior parietal cortex, which is adjacent to the PRR (see [54]).

But how does the brain compute these coordinate transformations? In the classical view (coming from geometry and applied in robotics, for instance), coordinate transformations are computed explicitly and applied sequentially; for example to pass from eye FoR to hand FoR, the brain would compute all the required transformations in series: between the eyes and the head, then between the head and the torso, and finally between the torso and the hand. We will see examples of this approach throughout the robotic part of the paper, in particular in Section III-C. However, in a novel view, coordinate transformations can be computed implicitly and in parallel [55]. The above mentioned neurons which encode for the difference between the hand and the target are a good example of such a view. In this particular case the only modality used for the coordinate transformation is vision; the positions of the hand and the target are
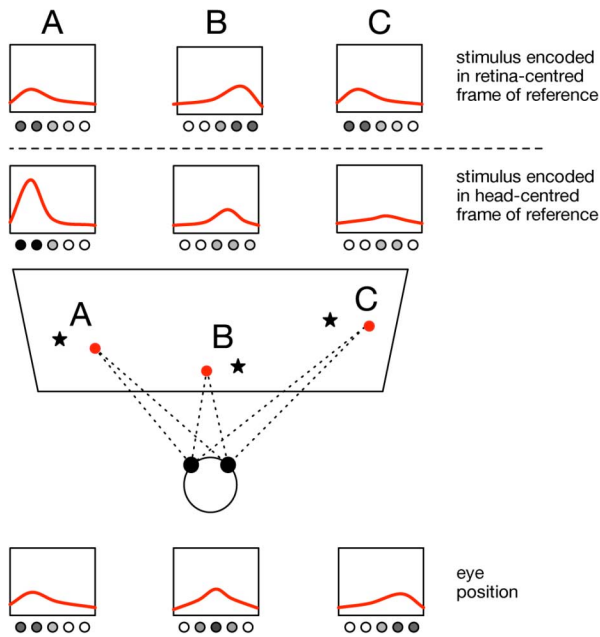
Fig. 2. Population-based encoding combined with gain fields to achieve coordinate transformations between a retina-centered FoR and a head-centered FoR (see text for details).

both acquired from the visual input. In fact, relatively little is known about the influence of proprioception for computing coordinate transformations.

One of the most relevant findings in brain research on coordinate transformations is that of gain modulation (also called gain fields, or fields of gain). Gain modulation consists of "a change in the response amplitude of a neuron that is not accompanied by a modification of response selectivity" [56]. It is a nonlinear way of combining information from two or more sources, let them be sensor, motor, or cognitive. Typically gain fields are used within a population-based encoding, in which several neurons respond to a region of space. The use of a population based encoding combined with gain fields for coordinate transformations is depicted in Fig. 2. The plots show the reconstruction of the signal obtained from different population of five neurons (circles below the plots). In situation A, the eyes are looking to the left (fixation point is marked by red filled circle) and a stimulus (a star) is further to the left. In the neural populations, it has the following effect: in the neurons that encode stimulus in the retinal frame of reference (top), we see more activation on the left; the neurons coding for eye position (bottom) also display higher activation on the left side; the population that has a head-centered reference frame encoding of the stimulus (middle) displays a combination (in fact a multiplication [57]) of the retina-centered stimulus and the eye position, as represented by a high amplitude in the left of the spectrum. The situations in B and C arise by applying the same rule. In C, for instance, the stimulus in the retina encoding is still on the left, but the eye is looking to the right. The head-centered encoding is a combination of these—in this case "contradictory" stimuli, with a mean in between and a lower amplitude. For instance, the parts of LIP and VIP discussed above (with position of a visual target in both eye-centered and head-centered coordinates) have

gain fields that depend on gaze direction, leading to body-centered coordinates useful for gaze control and object reaching [47], [58].

E. Body Schema and Forward Models

While the taxonomies (Section II-A) help us to roughly define the landscape of body representations, they still stop short off a concrete enough characterization that would allow one to build a computational or robotic model. The goal of this section is to illustrate some of the concepts put forth in the previous section on a simplified, but concrete enough, scenario and to clarify the relationship of body schema to the closely related concepts of peripersonal space and forward models.

Fig. 3 presents a didactic biologically motivated scenario, where an agent is interacting with an object on a table. The top part of the figure shows the agent and its visual field on the right, and a corresponding hypothetical neural architecture on the left. We have included three modalities: visual, proprioceptive, and tactile. In the visual modality, there are two hypothetical neural ensembles: one corresponding to the image on the camera (retina), and another which represents the same image in a body-centered reference frame. The position of the object as well as the hand is displayed in the activation. The transformation to the body-centered view is achieved by combining the camera image with the position of the head. Whereas regarding the position of the object, there is only visual information available, the location of the hand can also be obtained from proprioception in arm and hand.

Where is the body schema in this schematics? We offer the following interpretation: the activations in the individual neural fields are the *short-term body schema*—they represent the position or configuration of the body at one particular instant. The links between the neuronal ensembles, on the other hand, belong to a *long-term body schema*.[3] They are relationships between modalities that hold, at least over the here-and-now time-scale, and that can be used to perform coordinate transformations and to combine redundant information, such as regarding position of one's hand, in an optimal fashion.

Until now, only sensory modalities were involved. However, to move from one configuration to another, a motor action is required. A particular activation of arm and hand muscles can bring the agent to the situation at the bottom right part of Fig. 3, where it has moved the object in front of its single eye. Can the agent also learn about the mapping from the initial to the current state? The motor modality has to enter our representation, bringing us to the concepts of forward and inverse models (e.g., [59] and [60]). These concepts come originally from control theory but were adopted by the field of human motor control. Given the current sensory state and the motor command (or its copy—efference copy), a *forward model* can predict the next sensory state (or predicted sensory feedback—corollary discharge, bottom-left in Fig. 3). The so-called *inverse model* is a mapping in the opposite direction. Given a target (goal) state, and the current state, this model provides the motor command

---

[3]For reasons of simplicity, there are direct cross-modal mappings in our scenario. However, multimodal neural ensembles, i.e., those that fuse multiple modalities, were also reported in the brain. Similar connectivity could nevertheless apply to them as well.
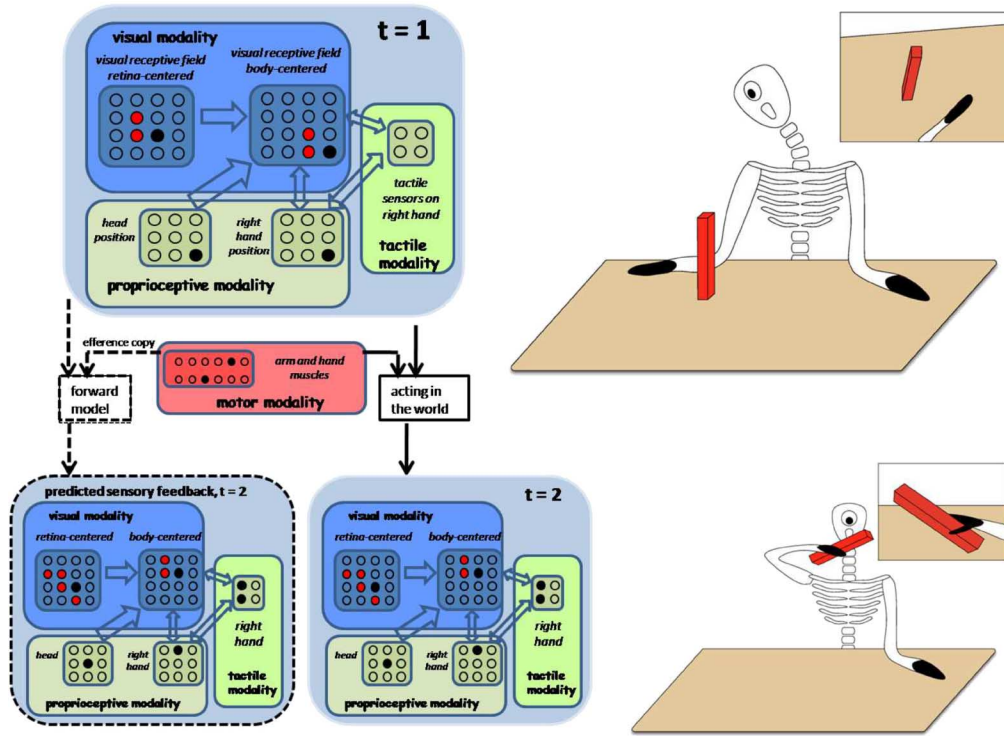
Fig. 3.   Long-term and short-term body schema, and forward models. This figure presents a simple biologically inspired scenario to illustrate the concepts. An agent is depicted on the right at two different time steps, while trying to grasp an object. The small window in the top-right of each picture shows the view from the agent's single eye (or camera). The left part depicts hypothetical neuronal ensembles of different modalities, their connections, and activations. Let us look at the initial situation (top). The agent is looking to the right and sees a red object in the center of its visual field (red activation in the retina-centered neurons) and its hand slightly right from it (black activation). However, with respect to the agent's body, the object is to the right and down. The retina-centered visual neurons can be combined with proprioception from the head muscles (black activation in the head position neurons) to perform a frame of reference (or coordinate) transformation, resulting in the activations in the body-centered visual neurons. Regarding the position of the hand, there is additional information from proprioception in the hand. The coordinate transformation between the visual modality (where the hand is seen) and the proprioceptive can also be performed and the two sources of information can be combined (double arrow between them). The bottom right part of the figure depicts a situation where the agent has grasped the object and moved it in front of its eye. The corresponding activations in the modalities are updated (object on the retina is much bigger, head and arm have moved) and there is a new activation in the tactile modality—on the agent's palm. The bottom left part illustrates the concept of a forward model: based on the multimodal map at $t = 1$ and a copy of a motor command, a prediction of the sensory map at $t = 2$ arises.

needed to reach the goal state. *Peripersonal space* can finally be also identified in our schematics. It is the space within reach; therefore, one possible instantiation would be part of the visual space, for which we can find a motor command (using the inverse model) to reach that space.

Forward models bring several advantages. For instance, the predicted sensory signals can be delivered before the real ones and can be exploited for control, or they can be compared with the real reafference and integrated to give a more reliable state estimation, or used to separate the expected effect of the agent's actions from unexpected intervention from the environment. As Grush [61] points out, the forward or inverse model (Grush uses the term emulator) can be either a look-up table storing previous input-output sequences, or it can be an *articulated model*—a model that includes some variables corresponding to their counterparts in the musculoskeletal system (e.g., elbow angle, arm angular inertia, and tension on quadriceps). Some of these variables can be measured (e.g., by stretch receptors) and these sensors can also be simulated in the emulator. Marques and Holland [62] propose to call the model that produces imagined sensory states more or less directly *unmediated*, and a model that produces them using a more or less complex self-simulation interacting with a simulated world *mediated*.

It should be clear by now that a body schema involves relationships between sensory modalities (such as coordinate transformations or integration of redundant information from modalities) and relationships between sensory and motor modalities. In our didactic scenario, these two components were separated—cross-modal mappings were between sensory modalities, whereas a separate forward model was dealing with a mapping between two sensory states, given a motor action. Such a division is very tempting and convenient for a robotic implementation. However, the biological reality may be more complex and it may not be possible to dissect the sensorimotor loop like this (see e.g., O'Regan and Noe [63] for a detailed account). Finally, the body schema has to involve not only spatial, but also biomechanical information, and it has to be plastic over time.

III. IMPROVING ROBOT BEHAVIOR THROUGH A BODY SCHEMA

Like natural agents, artificial agents can also acquire sensorimotor representations of their own bodies and use them to guide actions. However, before we discuss the character that such a model can have, we will discuss whether models or representations are in fact needed altogether. After, we will classify the various forms that a body schema of a robot can take.

We have used the nature of representation (explicit versus implicit) as a primary axis to divide the spectrum of body representations. Regardless of the representation, the key issue will be automatic acquisition and adaptation of a body schema. In particular, the application scenarios will include: recognition of own body, acquisition of its model, and its extension or adaptation when using a tool or after failure.

A. Does a Robot Need a Model?

The necessity for models of the world as well as of the robot itself comes as natural both to followers of traditional artificial intelligence or "good old-fashioned artificial intelligence" (GOFAI) [64], as well as to control engineers. The former work with symbolic models of the robot and the world, the latter use typically analytical models of the controlled system—plant models. This stance—that models, or representations, are necessary to produce useful behaviors—was challenged by the so-called new AI, behavior-based AI, or embodied cognitive science [65], [66]. New AI demonstrated the potential of robots that do not rely on representations, but rather on embodiment, and that exploit the interaction with the environment [67]. Relating back to our topic and paraphrasing Brooks, to what extent can it hold that "the body is its own best model?"

1) Intelligence Without Representation: It has been shown by the proponents of behavior-based AI that many remarkable behaviors can be achieved without a model. Examples are the achievements of Grey Walter [68], and Valentino Braitenberg [69] with purely reactive agents—agents that have no internal states, but only direct connections between sensors and motors. Another case in point that illustrates that a lot can be achieved without representation is the subsumption architecture of Rodney Brooks [70], [71]. Inspired by biological evolution, Brooks created a decentralized control architecture consisting of different layers. Every layer is a more or less simple coupling of sensors to motors (responsible for obstacle avoidance, for instance). Though in this architecture the individual modules may have internal states (as they are Finite State Machines), Brooks argues against even implicit representation here [65]. The "insect" robot Ghenghis [71] or the control architectures used by Cruse [72] demonstrate how a reflex-like controller can give rise to a walking pattern. There is no plan or model for the behavior in the robots' control architectures—walking arises only through the interaction of the body with the environment and simple sensor–actuator connections.

2) Model—Benefits and Costs: Before we ask ourselves the question, what is the best body representation for a particular robot, following up on the previous section, we propose to ask another question first: what are the benefits and costs of having a model of the robot's body?

The benefit number one typically is that the model of a robot (or plant) can be used for control. For instance, while multi-DOF robotic manipulators can be precisely controlled using the models and associated control techniques that were developed [73], to our knowledge, it is not feasible to control such a plant without a representation of any sort. With a precise representation of the robot's body—its kinematics and dynamics, including the actuation mechanism—it can be used for precise feedforward control with little or no feedback.

In controlled environments, such as industrial settings, this is sufficient. If feedback is present, the mappings from motors to sensors can also be learned, giving rise to a forward model (see Section II-E). Such a model can also be used to improve closed-loop control: sensory feedback can be predicted in advance—before it is actually received—and control action can be adapted (see, e.g., [74]). This is especially useful when the feedback comes with a significant delay. The fact that the expected feedback can be predicted can be also used to distinguish self-generated sensory information from sensory input generated by the environment. An account of similar scenarios in insects is provided by Webb [75]. A more elaborate and decoupled forward model, i.e., a model that can be iterated without actually executing the motor actions in the real world, can be used for planning of whole action sequences. Based on the predicted consequences, an appropriate action can be selected (e.g., [62]). As the last benefit, if the model includes a temporal dimension and uncertainty, using probabilistic terminology, it can be used to perform not only prediction, but also filtering (computing the belief state—posterior distribution over the current state).

However, we should not forget that there are costs associated with having models or representations. Such a model needs to be developed and that has costs attached to it. Heikkonen and Koikkalainen [76] report that robot programming—a substantial part of which is the development of the robot's model—accounts for about one third of the cost of an industrial robot system. The model is developed by engineers and given to the robot. This may be acceptable if the job has to be done only once—before the robot is put in operation. However, problems arise if the conditions change over time; this can be due to deformations of body parts from wear and tear, but it can also be due to more dramatic changes such as change of topology of the robot or the robot using a tool. In such situations, a significant part of the model would have to be reprogrammed giving rise to additional costs—model maintenance costs. This motivates the research in automatic model acquisition and adaptation.

B. What is Body Schema in a Robot?

It seems that in order for a robot to be able to perform a goal-directed action, two components are essential. First, to perform the action itself, it is often necessary to know at least some of the parameters of the system to be controlled. Second, if the robot relies on its own sensory system and if the goal is expressed in one of the sensory modalities (such as an object to be grasped in sight of a camera), a mapping between the sensory and motor modalities has to exist [77]. These two components can be almost completely separated or they can be completely intertwined. In robotics and control theory, the separation is typically clear. Even in the biological realm, there are indications that sensorimotor representations operate on kinematic variables, while the details that are necessary to perform a particular movement (an inverse dynamics model of the "plant" which needs to include inertia, stiffness, possibly actuator dynamics, etc.) can be delegated to other control structures (such as cerebellum and the spinal cord [78]) and to the body itself.

Let us first look at a prominent scenario, a multi-DOF robotic manipulator. The typical goal is to make the end-effector reach

TABLE II
BODY SCHEMA TO IMPROVE ROBOT BEHAVIOR—AN OVERVIEW

Study	Key issue	Platform (S:simulated, R:real)	Body representation
Hersch *et al.* [82]	automatic kinematic chain calibration	humanoid robot (S,R)	kinematic chain
Martinez-Cantin *et al.* [94]	automatic kinematic chain calibration	humanoid torso (S,R)	kinematic chain
Nabeshima *et al.* [95]	body schema extension	humanoid torso (R)	kinematic chain
Bongard *et al.* [97]	body schema acquisition and adaptation	quadrupedal robot (R)	kinematic chain
Sturm *et al.* [83]	body schema acquisition, adaptation and extension	robot arm (S,R)	Bayesian network
Metta *et al.* [99]	body schema development	humanoid torso (R)	look-up table
Gaskett and Cheng [100]	body schema acquisition and adaptation	humanoid torso (R)	self-organizing map
Yoshikawa *et al.* [101]	self-recognition, body image acquisition	humanoid torso (R)	cross-modal map
Gold and Scassellati [102]	self-recognition, agency	humanoid torso (R)	Dynamic Bayesian Network
Natale *et al.* [103]	self-recognition, body schema development	humanoid torso (R)	mutli-layer perceptron
Dearden & Demiris [104]	body schema acquisition	camera and grippers (R)	Dynamic Bayesian Network
Grimes *et al.* [105]	body schema acquisition	walking humanoid (S,R)	Dynamic Bayesian Network

a certain point in the workspace. While the goal is typically expressed in Cartesian or visual space, motor commands will be issued in joint space. Thus, a coordinate transformation between the two spaces is essential (confront with Section II-D and with the notion of peripersonal space in Section II-C3). An example of such a mapping is inverse kinematics, i.e., the manipulator joint angles needed to achieve the desired position and orientation of the end-effector in Cartesian space can be obtained. In industrial settings, a manipulator can often operate based solely on the kinematic model, without visual feedback. The dynamics (forces/torques needed to achieve desired positions) can be delegated to another subsystem (e.g., feedback controllers within servo motors), or a separate dynamical model of the plant can complement the kinematic model. We call the model used in this case *explicit*. The kinematics (and dynamics) are described by equations; the parameters, such as segment lengths and orientation of joints, are measured and inserted into the equations. The platform and its model are then carefully calibrated. We can call the model also *objective*; an attempt is made to objectively measure the physical reality of the robot and input it in the model. Yet, we are dealing with a representation of the robot's body that can be used to guide actions, and thus, it can be classified as a body schema. For us, however, it will lie on one end of the spectrum of research and we will discuss it only briefly. First, because we feel that such a model departs too far from the properties that we attribute to a body schema; contrary to its biological counterpart, this model is typically fixed, explicit, precise, and centralized. Perhaps even more importantly, it involves minimal or no perception; it is given from the outside and thus relies on information that biological agents cannot access. Second, modeling and control of robotic manipulators (e.g., [73], [79], and [80]), or robots in general (e.g., [81]), is already an enormous research field in itself.

Articulated models come closer to the notion of body schema as we know it from biology. Recall from Section II-E that an *articulated model* is based on state variables (such as manipulator joint angle positions) that interact according to the laws of dynamics and mechanics [61]. This time, however, the variables have to be measured by the robot's own sensors. Hersch *et al.* [82] hence use the term *subjective body schema*. Usually, the definition of state variables comes from the outside with prior knowledge of the problem. The model can still have a form of equations, as in [82]. However, we will regard even a body schema that does not have a mathematical form as explicit, if

there is a one-to-one correspondence between the body parts of the real robot and those in the model, as in [83]. Articulated models will be discussed together with explicit models.

Explicit models have a number of advantages. The sensorimotor mappings, as well as plant models are governed by explicit equations, and hence it is possible to calculate the behavior of the system even in previously unseen situations. Also, as they are more transparent, it may be easier to debug them and to assess their performance. However, as the plant and sensorimotor mappings become nonlinear (imagine a compliant pneumatically driven robot with multiple modalities), a closed-form solution may not exist. Platforms that cannot be modeled explicitly will be addressed by *implicit models*. Such a body representation can be a simple look-up table with previously encountered sensorimotor relationships, or, neural networks often serve as the substrate for an implicit body schema. These models are typically more bio-inspired and will close the section on improving robot behavior through a body schema. At the same time, they will provide a natural transition to Section IV—that deals with robots as tools to model biological body representations.

The representations of the robot's body, as discussed above, contain the long-term properties of the plant and hence correspond to the notion of a long-term body schema (cf. Sections II-A and II-E). However, what is no less important is a short-term representation of the body—where it is in space right now, for instance. Current sensory readings have to be mapped onto some states (if there are states) in the long-term body schema and can then be used to plan future actions, for instance. The short-term body reprepresentations can have a "winner-take-all" form, or they can have a probabilistic form, where alternative states are possible, with given probabilities (cf. gain fields and population based encoding in Section II-D).

The most prominent studies—using both explicit and implicit representations—that we will review are summarized in Table II.

C. Explicit Models

Fixed kinematic models will start off the section concerned with explicit representations of robot bodies. Then we will move to adaptive models—models that can self-calibrate or that can even learn the topology of the body structure. These models are inferred using the robot's own sensors and hence are subjective, even though the perception is typically simplified. Finally,

we will discuss models of the robot's body that also include dynamics.

1) Fixed Kinematic Models: Let us briefly look at a multi-DOF robotic manipulator again. It operates based on its forward and inverse kinematic functions that ensure the coordinate transformation between the workspace (a Cartesian coordinate system in which the goal for the end-effector is expressed), and the joint space. The joint positions can be directly used as target commands for servomotors.[4] If the manipulator is accompanied with a fixed camera that is observing the environment, an additional frame of reference transformation from the camera frame reference to the Cartesian or task space has to be defined.

What are the limitations of this architecture? First, a fixed kinematic model applies to robots obeying rigid body dynamics only. Second, the model is designed from the outside and is not adaptive. New calibrations have to be done in response to plant drift (e.g., robot's wear and tear). A change in the robot's geometry or the addition of a tool might require a new model. Third, this approach is not easily extensible to include more modalities (such as touch). Additional nonlinear sensorimotor mappings and their integration cannot be dealt with by the current analytical machinery. Fourth, since dynamics was not addressed by the kinematic model, this solution has variable performance in different tasks, where the end-effector has to apply force, or when external forces such as gravity loading change, or with plants that cannot be directly position controlled (e.g., pneumatic actuators). All these shortcomings will be addressed in the following sections.

2) Self-Calibration of a Parametrized Kinematic Model: Self-calibration of a parametrized kinematic chain can deal with changes in geometry over time (such as changes due to material fatigue). Automatic calibration is only possible when the system receives information from more than one source. For instance, the calibration of a camera-manipulator system can be achieved automatically by comparing the position of an end-effector as observed by the camera with the one from the forward kinematic function (after they have been converted to a common, typically Cartesian, reference system). Leaving the human engineer out of the loop can reduce costs. As a special case of self-calibration, we include body schema extension in this subsection. Automatic calibration of a model is addressed by some traditional methods from machine learning [84], system identification [85], [86], or probability theory [87]. More specifically, there is a number of solutions to the automated calibration of a kinematic chain [88]–[90] or a hand/eye setup [91]. Typically, a sampling period in which different configurations are visited is followed by an optimization procedure.[5] However, rather than "batch adaptation," it is desirable to develop systems that learn incrementally and online, following the inspiration from biology.

Hersch *et al.* [82] present an extension of the self-calibration approach. Taking advantage of prior knowledge of its kinematic

(a)

(b)

Fig. 4. Robot and its body schemata. (a) Hoap3 robot. (b) Body schema. Left: "Real" schema. Middle: body schema learned by looking at hands and feet only. Right: Body schema learned when looking at additional joints. Hersch *et al.* [82].

structure (number, arrangement and type of DOFs), a simulated 24 DOF humanoid robot is able to learn the missing parameters of the kinematic chain—position and orientations of joints—by observing its body with a camera. A gradient descent algorithm is then applied, the efficiency of which increases when additional joints, not only the end-effector, can be observed. On a real robot (Hoap3, see Fig. 4), it was demonstrated that the algorithm can cope with the incorporation of a stick as an extension to the body within two or three minutes (cf. Section II-C3). In [93], this system is complemented by learning the neck-eyes kinematic chain using optical flow, and the whole system is demonstrated on the iCub humanoid robot. There are a couple of features that bring this work closer to the biological notion of body representations. First, contrary to the standard calibration approaches in which a phase of sampling and optimization precedes the actual use, the algorithm of Hersch *et al.* [82] works online. Second, it is a case of a "subjective body schema." The system is self-contained, or situated, in the sense that the sensorimotor mappings learned are solely based on the information acquired from the robot's own sensory and motor signals. The geometrical properties of the robot, such as the segment lengths, of course mediate the sensorimotor relationships, but cannot be accessed directly. The correspondence between the different reference frames, e.g., from end-effector to head with the camera, is given by the kinematic chain parameters which are subject to learning. Thus, there is no precoded transformation given from the outside, such as one from a camera to Cartesian frame.

Martinez-Cantin *et al.* [94] presented an improvement in efficiency over the work of Hersch *et al.* First, they employed a more efficient learning method than gradient descent for estimating the body schema parameters: a recursive least squares (RLS) estimator. Second, they explored the configuration space

---

[4]This is often the case in robotics: proprioception from joints can at the same time act as a motor command—it is the target position sent to a servomotor. However, although this simplification may be convenient, we have to be aware that it departs from biological reality.

[5]This is not the case for the exploration-estimation algorithm [92] though, where the exploration strategy is more sophisticated and intertwined with the model evaluation stage.
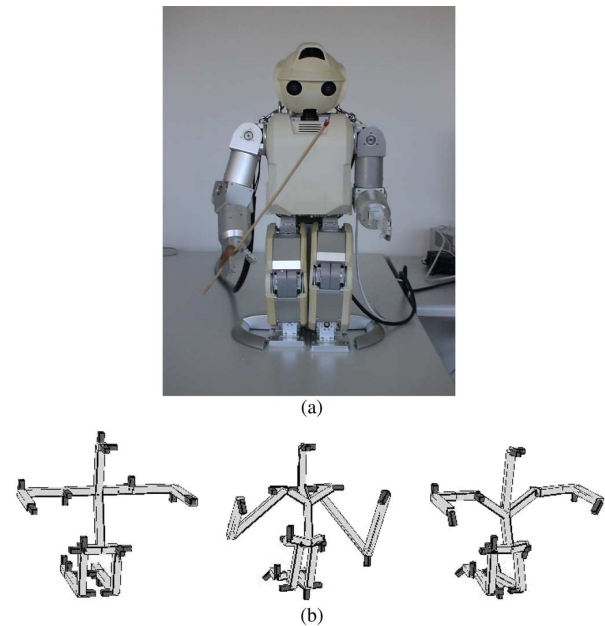
in an intelligent way, looking for the most informative measurements based on the posterior uncertainty from the RLS.

Nabeshima *et al.* [95] also employ a traditional kinematic controller. However, the problem they address is not self-calibration, but specifically body schema extension and the detection of such a change. An upper humanoid torso is used to reach for objects. Apart from proprioception (joint angles) and vision, a third modality, touch, is involved. When the robot hand touches a target, a learning process—spatio–temporal integration of the multimodal information that preceded the contact—is triggered. This can be retrieved later from an associative memory and used to drive a controller. When the robot arm is extended with a stick (a primitive tool), contacts occur in new situations, and a new kinematic controller is learned in response. Neural networks are employed to implement the spatiotemporal integration and learning. This work is much more bio-inspired than what we have encountered and will encounter in this section. However, as no explicit correspondence with biology is established, we do not classify the work as biological modeling (an example of which, Hikita *et al.* [96], will be presented in Section IV).

3) Automatic Model Synthesis Including Topology: In this section, we review robot kinematic models that can be synthesized automatically with little prior knowledge. Contrary to the previous section, no parametrized form of the model is necessary. As a result of that, not only parameters like segment lengths, but also the robot's topology can be learned. Therefore the work reviewed in this section does not only address body schema extension, but can cope with more dramatic changes in the robot's body, such as the loss of a limb or a blocked joint, leading to resilient machines. We will focus on two case studies: 1) the work by Sturm *et al.* [83], who show how a robotic manipulator can synthesize and adapt its kinematic model from self-observation and can then use it for reaching; and 2) the work by Bongard *et al.* [97], in which a quadrupedal robot continuously models itself and generates new locomotion patterns.

Let us first point out what the two models have in common. First, both models are explicit in the sense that there is a one-to-one match between the components (e.g., body parts) in the body schema (or model) and their counterparts in the physical robot. The number of joints and body parts presents the prior knowledge. Second, the controllers operate on kinematics (dynamic disturbances are handled by position-controlled servo motors), and only static configurations (i.e., not the dynamics of behavior to reach that configuration) are used to assess the match between the model and the physical robot. Third, both present a case for a "subjective" schema, as the signals from the robots' own sensors are used to validate the model. And fourth and last, there is a population of candidate models involved.

Let us start with the work of Sturm *et al.* [83]. Here, different robotic manipulators are used (4, 6, and 7 DOF). The robot observes the pose of its body parts (with special visual markers) using an external monocular camera (see Fig. 5). The goal is that the model of the manipulator is learned through exploratory actions and self-observation. In Hersch *et al.* [82], described previously, the parameters of the kinematic chain were learned, providing a coordinate transformation between two sensory modalities—visual (camera) and proprioceptive (joint space). On top
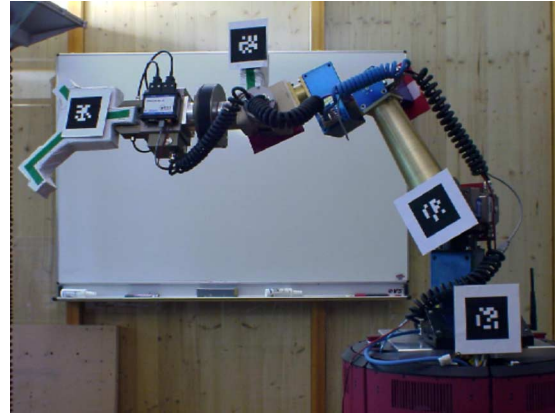

Fig. 5. A 6-DOF robotic manipulator arm learns and monitors its own body schema using an external monocular camera and visual markers. Sturm *et al.* [83].

of that, different, also classical, approaches to control can be used, and will have to provide a mapping between motor commands and joint angles. Thus, although their body representation could be used for action, it does not contain the motor modality directly. Sturm *et al.*, on the other hand, directly include the action commands. As we will see, their architecture thus also provides a forward and inverse model of the robot (cf. Section II-E).

The body schema in Sturm *et al.* is the joint probability distribution of available actions signals (target angles sent to individual joints), self-observations (as obtained from the camera), and true poses of the body parts (hidden states). The body schema is modeled as a Bayesian network, where the nodes correspond to body parts, action signals and model components. The structure of the network reflects the kinematic chain. For example, the 6-D pose of a body part of the manipulator depends on the pose of its predecessor and one of the action signals. These dependencies enter the Bayesian network. The learning problem is then factorized into two parts: First, local models that describe the relationship between pairs of body parts are learned using Gaussian processes regression. Local models that do not explain data well are discarded. Second, a graph is built from the valid local models. Under the assumption that the manipulator has no cycles, the problem of finding the kinematic structure of the manipulator corresponds to the minimum spanning tree of this graph. The cost function is defined as the combination of the marginal data likelihood and a complexity penalty for each local model. Instead of using joint encoders, the relationship between motor commands and positions of the body parts of the manipulator is learned directly, circumventing the mapping between the target motor commands and the angle actually assumed by the joints. In order to control the manipulator, an inverse model is needed, i.e., a mapping from desired pose to action commands. While this can be obtained by searching for the motor commands that maximize the likelihood of generating the desired pose, it results in a high-dimensional optimization problem. Therefore, a different approach is used: the representation of the model allows to apply differential kinematics, in particular, it is possible to compute the Jacobian of the forward model and thus a
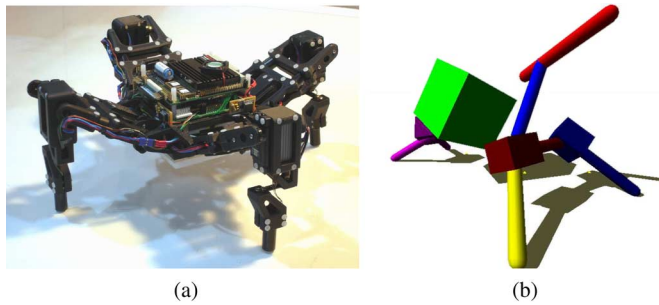
(a)   (b)

Fig. 6. Robot and its (incorrect) candidate model. Bongard *et al.* [97].

gradient-descent algorithm is used for selecting suitable motor actions.

In their experiments, Sturm *et al.* demonstrate that: 1) the robot can learn its kinematic model from scratch; and 2) the robot can adapt the model to blocked joints as well as to deformations. This presents a solution to automatic model synthesis, calibration, and body extension, as well as recovery from damage. Furthermore, the model provides additional benefits thanks to its probabilistic nature. First, information from the robot model is combined with the sensory data in a statistically optimal fashion, and the model also contains uncertainty of the estimates. Second, each model candidate has an associated likelihood, and thus, multiple candidate models explaining data can be kept in parallel. Classical control, which assumes a single model, can thus be extended to take the uncertainty into account. Third, extending the model in time would allow to perform prediction, or filtering (computing the belief state). Therefore, the Bayesian framework encompasses both long-term (structure and parameters of the network) and short-term body schema (current belief), and a forward and inverse model, including a measure of the reliability of the information.

Bongard *et al.* [97] used a different platform, a quadrupedal robot, whose body schema is to serve the synthesis of locomotor behaviors. Compared to the manipulator arm scenario, the interaction with the environment is much more profound here. A model of the dynamics (mass and inertia) of the robot, as well as of the ground and their interaction (friction model) is indispensable. The robot's self-model is split into two parts here. The first part consists of an externally designed model of the robot and the environment in a physics-based simulator. This is a special form of an explicit model—equations of motion for the mobile robot are not specified analytically, but they are embedded in the physics-based simulator and numerically integrated. This first part of the model contains the robot as a chain of rigid bodies connected by servomotors, and remains fixed during experiments (is "known" to the robot). The second part is the kinematic structure of the robot in the simulator, i.e., how are the rigid bodies connected. This part is unknown to the robot and is subject to learning and adaptation. Fig. 6 shows the real robot and one candidate model (with incorrect kinematic structure) in the simulator. To validate the model the information obtained from the sensors on the real robot is compared with the one from the simulated sensors in the simulator (cf. with the notion of emulator and articulated model of Grush [61] and with mediated model of Marques and Holland [62] in Section II-E).

A set of 15 candidate self-models is kept. In every (static) configuration of the robot, the sensor readings are taken and compared with the readings from the simulated robot in the same configuration. In [97], only orientation sensors are used, but in [98], more modalities are employed and their relevance is also assessed. However, these configurations—or action that lead to them—are not selected at random. It is the action that is expected to best disambiguate between the candidate models that is executed on the real robot. Behavior synthesis on the model pool thus precedes and only when the information expected to gain is maximum, actions are executed on the real robot. On damage (lower leg part breaks off), a mismatch between the predicted and real sensory signals is detected, and exploration, modeling, and testing is reinitiated until a new model which reflects the change is found.

This architecture also encompasses a forward model. Whole action sequences can be executed in the simulator, and their outcomes observed. New behaviors can thus be synthesized in the model first, which would otherwise be a lengthy process on the real robot. This is an advantage of an explicit model. Unlike an implicit model, which allows to interpolate between actuation-sensation patterns that have been seen before, an explicit model allows to extrapolate, and to generate qualitatively new behavioral patterns. Nevertheless, the (explicit) interaction with the environment is very hard to model in this case (e.g., contact modeling with the ground is a notorious problem) and there is always going to be some discrepancy between the model and reality (the reality gap). Although the parameters describing this interaction (such as friction) were fixed in the physics-based simulator and represented prior knowledge in the cited work, in principle, they could also be adaptive.

4) Models Including Dynamics: Apart from the physics-based simulation used by Bongard *et al.* [97], so far we have dealt with kinematic models only—the forces and torques required to cause a particular motion were not addressed. Nonetheless, these are essential to finally execute an action. This gap is filled by (inverse) dynamics models of the robot or plant. This can be viewed as a relatively independent module and there are indications that a similar strategy is used in biological motor control [59]. Therefore, the models of robot dynamics do not lie at the center of our interest and we refer the reader to many textbooks on the topic, e.g., [73], [79], and [80].

Having drawn this parallel between inverse dynamics in biological and robotic motor control, let us also point out the important differences between them. It is probably fair to say that the basis of the field of control in robotics is largely formed by engineered models (e.g., computed-torque control [73], [79]). While model adaptation and dealing with uncertainty is also addressed (by robust and adaptive control [79]), adaptation to dramatic changes in the robot dynamics lies still outside of the scope of these methods. Similarly, the platforms that can be modeled are still largely restricted (mostly stiff rigid bodies). On the other hand, we know that biological motor control can deal with both significant changes to the dynamics or to the kinematics, and with compliant platforms, for instance. Therefore, if we want to deal with such robotic platforms, we may need to resort to implicit models, and this takes us to the next section.

D. Implicit Models

This section reviews work where an implicit representation of the robot's body is used. This can take a form of a simple look-up table or it can be a neural network, for instance. We will also review work that deals with self-recognition—how does the robot find its body and separate it from the environment. Finally, we will look at models that address the issue of delays in the effects of robot's actions. Compared to the explicit models, much less prior knowledge enters the implicit representations.

1) Representations of Sensorimotor Mappings: To derive analytical equations representing the kinematics and dynamics of a controlled system is not always possible. In situations where this is not feasible (in highly nonlinear systems with compliant actuation composed of deformable bodies, for instance), these mappings can still be learned using different function approximation techniques. Such mappings can either aid standard control schemes (as in the case of neural networks for control), or they can be control schemes in their own respect.

If a model of a plant cannot be obtained analytically, it is still valuable to obtain a model that treats the target system as a black-box. Its input–output behavior can be learned by a system identification process. By observing the responses of a system to different inputs a forward model can be learned. For control, however, an inverse model is typically required. This can be either obtained by inverting the forward model (which is possible only in special cases), by directly learning the inverse mapping, or by the so-called distal supervised learning approach [106]. *Inverse kinematics* can be approximated by various approaches: locally weighted regression [107], multilayer perceptrons, or radial basis functions [108], [109]. Over the past decades, connectionist approaches have been integrated into numerous control architectures (for instance model reference adaptive control, model predictive control, internal model control [110]), where they form one or more of the building blocks: plant model, inverse plant model, or controller. One of the earliest architectures that is still being developed is the cerebellar model articulation controller (CMAC) [111], [112]. We will refer the reader to the abundant literature on the topic of neural networks in traditional control schemes [108], [113], [114]. Interestingly, unsupervised (or self-supervised) neural network architectures can also be used. Barreto *et al.* [115] demonstrate the use of self-organizing maps and some of their advantages. For instance, the topological arrangement of network nodes ensures that a redundant manipulator is well-behaved. A "lazy" cost function is implicitly coded—while looking for an adjacent target point, an adjacent joint configuration is automatically selected.

The big advantage of implicit approaches is that almost arbitrary sensorimotor mappings can be represented. For instance, *inverse dynamics* does not present a problem with different characteristics, assumptions, and complexity than inverse kinematics, as is the case with explicit modeling. If dynamic instead of kinematic variables are fed to the learning algorithm, inverse dynamics can be learned in a similar manner (e.g., [116]). Similarly, platforms that were outside of the scope of analytical modeling, such as pneumatically driven robots, can now be treated equally easily [117]. The problems of coordinate transformations and forward modeling do not have to be addressed as separate building blocks anymore.

To further illustrate the case of sensorimotor mappings, let us look at visually guided reaching. This is a hand–eye coordination problem and there are two basic strategies to tackle it: 1) open-loop control, in which a sensorimotor map that relates the hand visual location and the arm position from proprioception is needed; and 2) closed-loop control, where the visual Jacobian of the manipulator is needed. The open-loop strategy can be realized through a combination of classical explicit frame of reference transformations that involve the hand, body, and camera reference frames. As mentioned in Section III-C2, these maps can be obtained through automated calibration procedures (kinematic chain [88]–[90], hand/eye setup [91]). However, a highly structured environment is typically required for these calibration procedures (see [118] for more details). The Jacobian that is needed for the closed-loop strategy (or visual servoing [119]) can be derived analytically, or estimated [120].[6] The two strategies, open-loop and closed-loop can also be combined, as demonstrated by Natale *et al.* [118], for instance, where reaching in 3-D is possible without prior knowledge of the kinematic model.

The mappings needed to perform visually guided reaching can also be coded implicitly. For instance camera calibration and triangulation can be learned in an implicit manner [121], [122]. Moreover, interestingly, the open-loop component which requires a sensorimotor mapping can be turned into a motor–motor coordination problem, as demonstrated by [99], [100], [123]. Rather than learning the mapping between visual space and arm motor space directly, the eye–head system is exploited. A camera is let to fixate on the target (this can be precoded or learned separately) and the appropriate motor variables of the eye–head plant are extracted and used to learn the relationship with the hand motor plant variables that represent reaching to the target. This relationship can be represented by a look-up table [99] or by a self-organizing map [100]. The learned mapping reduces the dimensionality of the problem, and is an instance of a body schema which allows to reach to a certain point in space—the target to which the eyes are looking—in an open-loop fashion. Metta *et al.* [99] also spell out the important features that characterize their approach: 1) the kinematic and dynamic parameters are not explicitly identified as in classical control theory approaches; and 2) there is no distinction between the system's calibration and control. In other words, the two processes are completely intertwined, and the performance of the overall system can grow in an incremental fashion over time.

2) Self-Recognition: In the works that we have described so far, the goal was to acquire or adapt a body representation. The representation itself has taken various forms—an explicit kinematic chain, a model in a physics-based simulator, or a cluster of implicit sensorimotor mappings. However, it was assumed that a robot knows which signals come from its body. For instance, in the work of Hersch *et al.* [82] or Sturm *et al.* [83] (see Sections III-C2 and III-C3), all the body parts of interest were visible and easy to distinguish. In reality—if we take a developmental perspective and assume that the robot does not have

---

[6]The Jacobian is a good example that sensorimotor maps can represent relationships between higher order (in this case first-order since Jacobian is a derivative) variables as well.

this prior knowledge—the robot first needs to 'find itself' in the stream of sensorimotor signals (cf. Section II-C1).

Yoshikawa *et al.* [101] address the problem of how a robot identifies its arms in a visual image. Unlike objects in the environment, the arms remain at fixed positions, and due to this invariance, they can be extracted from the visual scene and identified as belonging to the body. Hebbian learning is employed to pick up this invariance between the visual modality (disparity after the eyes fixate on an object), and proprioception (position of cameras—pan, tilt). The work of Yoshikawa *et al.* [124] is an extension of this strategy to multiple visual attributes (disparity, luminance, chroma, and edges). Since the arms are not allowed to move, the procedure is dominated by perception and we can talk about *acquisition of body image* (cf. Section II-C1).

A largely converse strategy is employed by Fitzpatrick and Metta [125], Natale *et al.* [103], and Gold and Scassellati [102]. It is the active behavior of the robot that is used to self-recognize. Kemp and Edsinger [126] can perhaps be viewed as a transition between the two strategies. The robot's arms are allowed to move, but it is *spatial contingency*—mutual information between salient patches in the visual scene and expectations on appearance and position of the robot's parts—that allows self-recognition. On the other hand, it is *temporal contingency* that is utilized in [102], [103], [125]. The robot learns to recognize its body parts *because* they are moving. However, since external objects can be moving as well, it is the correlation between the visual input (optic flow) and the motor signal that facilitates the body identification [125]. Natale *et al.* [103] improve the robustness of this procedure by using periodic hand motions. Then, the robot's hand could be segmented by selecting, among the pixels that moved periodically, only those whose period matched that of the wrist joints. Gold and Scassellati [102] use probabilistic reasoning and examine the likelihood of three alternative models: 1) robot's own motors generated the movement; 2) something else generated the movement; or 3) irregular movement. Case (1) would correspond to the robot's own body. Unlike the case of Yoshikawa *et al.*, action plays a key part in these methods. Therefore, it is more appropriate to talk about *body schema acquisition* and *sense of agency* (cf. Section II-C1 again). We also want to point out that this strategy can be naturally extended to action recognition in others and imitation (see [102]), tool use, or interaction with objects (see [103]).

3) Temporal Models: We have seen how an agent can exploit temporal contingencies to self-recognize. However, once the agent has found its body, should the temporal domain be still preserved in the synthesis of body representations? The sensorimotor mappings that were discussed so far were largely relationships between various modalities in *static* configurations. Some architectures encompassed a forward or inverse model and thus, allowed to iterate a body state in time. However, in reality, different actuators as well as sensors have their specific time delays associated with them. A body schema unfolded in time can be nicely represented with a dynamic Bayesian network (DBN). Dearden and Demiris [104] used a similar approach to Sturm *et al.* [83], but included motor delays into the body schema. The problem of model selection among competing candidate body schemata (as represented by the DBN) has thus grown to include the temporal dimension. Hidden states are discrete and represent

the states of two grippers (open/closed). Observables are based on optic flow in the visual scene; visual blobs are extracted and clustered with a k-means algorithm. The prior knowledge that enters the body schema is the "template" for the structure of the Bayesian network: from motors to hidden states to observables. While this approach is more general than Sturm's [83], the toll that needs to be paid is that the system has much fewer DOF (essentially 2).

The work by Grimes *et al.* [105] uses a similar approach, but addresses a different problem: bipedal locomotion. Humanoid walking is a much more difficult problem than robotic manipulation. Balance becomes a key issue, modeling dynamics becomes inescapable, and we have to deal with a floating-base system. Traditionally, explicit modeling is performed based on CAD data, followed by further parameter estimation. The most famous control scheme in use is the zero-moment point (ZMP) control [127]. While this is commonly applied in walking humanoids ([128]–[130]), it has not yet been possible to extend it to rough terrain ([131] is an attempt in this direction, but on a quadruped platform). Therefore, Grimes *et al.* [105], instead of using an explicit physics-based model of the robot and a control scheme on top of this, adopted a model-free, or implicit, approach. The kinematic and dynamic states are represented in a DBN, together with action commands and observables. The problem of balance is addressed by a relationship between sensors (gyroscope and pressure sensors), which is, again, an instance of a subjective or situated body schema. Parameters for the model are learned with Gaussian processes. Implementations with Bayesian networks have the usual benefits that they allow for prediction, planning, or filtering, all that with measures of uncertainty. Moreover, both Dearden and Demiris [104] and Grimes *et al.* have shown how to utilize their architectures in an imitation scenario.

IV. ROBOTS AS MODELS OF BIOLOGICAL BODY REPRESENTATIONS

As we have seen in Section II, although direct recordings from the brain have revealed relevant facts about body representations in biology, the mechanisms underlying the working and the development of body schema (and body image) in animals and humans are still far from clear. A difficulty in understanding such mechanisms from the observation of neural activity alone is that it is hard to separate the influence of the target mechanism on the recorded data from a variety of other processes inside the brain, as they result from the interaction among brain, body, and environment. A synthetic approach—investigating the phenomena of interest by implementing them in robots (e.g., [132])—is a promising methodology to overcome the difficulties that computational neuroscience faces. Not only the mechanisms underlying a mature body schema, but also its development in infants can be addressed by synthetic modeling (Asada *et al.* [133] provide an excellent review). Body schema implementations that aim at modeling biology naturally feature more biologically realistic architectures and mechanisms. Hebbian learning, self-organizing map (SOM), or spike timing-dependent plasticity (STDP) are often employed. In some cases, it is possible to establish a correspondence between the proposed models and neural firing patterns in the cortex [96], [134]. While

TABLE III
MODELS OF BODY REPRESENTATIONS IN BIOLOGY—AN OVERVIEW

Study	Key issue	Platform (S:simulated, R:real)	Body representation
Yoshikawa *et al.* [136]	body image acquisition, multimodal representation	humanoid torso (R)	cross-modal map
Fuke *et al.* [135]	body image acquisition, extension to invisible body parts	humanoid torso (S)	self-organizing maps
Fuke *et al.* [134]	body image acquisition, coordinate transformations	humanoid torso (S)	self-organizing maps
Hikita *et al.* [96]	body image/schema acquisition and extension, self-recognition	humanoid torso (R)	self-organizing maps
Morasso & Sanguineti [138]	body schema acquisition	robotic arm (S)	self-organizing body schema
Stoytchev *et al.* [143], [139]	body schema acquisition and extension	robotic arm (S,R)	self-organizing body schema
Pitti *et al.* [140]	body schema acquisition, self-recognition, agency	robotic head (R)	spiking neural networks
Kuniyoshi & Sangawa [141]	low-level body representations	humanoid (S)	self-organizing maps
Marques *et al.* [144]	low-level body representations	humanoid (R)	–

it is probably fair to say that this body of research is at its nascent stage, there are a couple of relevant cases that will be described below. The scenarios we will come across will resemble the ones from Section III, but this time, the architectures will not merely draw inspiration from biological body representations, but will explicitly attempt to model the biological mechanisms.

We will structure this section as follows. Many synthetic studies have been carried out to understand multimodal body representations which, in primates, are found in the parietal cortex. Here, we categorize them into two groups: nonaction-oriented body representations (body image), and action-oriented ones (body schema). The former body of work employs cross-modal maps that are modified through Hebbian learning applied on individual modalities [96], [134]–[137]. The latter category comprises studies in which the acquired body representations are utilized to coordinate the robot's behavior [138]–[140]. Third, we will review the work by Kuniyoshi and Sangawa [141] where the emphasis is placed on the physical interaction between body and environment and on the effect of low-level (spinal) control. On top of these, low-level sensorimotor representations can emerge. The most prominent studies that we will review are summarized in Table III.

A. Nonaction-Oriented Body Representations (Body Image)

Many synthetic studies have focused on how to integrate information from tactile, visual, and proprioceptive sensor spaces. The "body maps" that are acquired are used for recognition of the agent's own body (cf. Section III-D2). Yoshikawa *et al.* [136] focused on correlations in the activation of tactile, visual, and proprioceptive modalities. Through an experience of self-touching, maps linking the modalities were associated by Hebbian learning. While Yoshikawa's study allows to represent only body parts that are visible to the robot, Fuke *et al.* [135] proposed a model in which the invisible parts—the robot's face—can also be incorporated into the body representation. This was done via learning a Jacobian from the motor (joint) space to the visual space. Integrating the velocity, position in visual space can be estimated for invisible parts as well. Then, while the robot was touching its face with the arm, the position in the visual modality could be estimated and matched with the touch modality—learning a cross-modal map. It is then hypothesized that a fetus establishes this correspondence while touching its face in the womb and this may explain why a newborn is able to respond to faces immediately after birth.

Another important topic is body schema adaptation during tool use (see Section II-C3). While we have encountered implementations of this behavior in the section on applications (e.g., [82], [83], and [95]), the mechanisms employed were only inspired by biology. The approach of Hikita *et al.* [96], on the other hand, models the mechanisms hypothesized to be used in humans. In particular, they focus on the role of the attention system in detecting body extension by a tool. Based on a neurobiological model by Itti *et al.* [142], a model that enables a robot to detect its own end-effector by associating proprioceptive information with visual information during visual attention, a saliency map, is proposed. Tactile sensation on the robot's hand is used to trigger the association.[7] The representation enables a real robot to recognize its own body and there is an analog to the findings in parietal cortex during use of a tool, as described in [27] (see Figs. 7 and 8).

Fuke *et al.* [134] extended the problem of integrating tactile, visual, and proprioceptive modalities by addressing the frame of reference transformation that needs to occur between an eye-centered and a head-centered reference frame. A model was proposed according to the relation between VIP and LIP in human and primate brains as described in Section II-D. Based on studies on infants [15], this integration is assumed to be achieved through hand regard behavior: human infants gaze at their own hands in front of their face at around four months of age. In the experiments of Fuke *et al.* a robot first acquires a head-centered visual space representation by associating ocular angles and camera images while gazing at its hand moving. Then, it integrates tactile sensations with visual stimuli by touching its face. Experimental results with a simulated human-like robot show that the activities of the acquired maps are similar to the ones of the VIP neurons as observed in [47]. The correspondence of the model with brain regions is shown in Fig. 9.

B. Action-Oriented Body Representations (Body Schema)

This section deals with models of biological body representations used to guide actions. In some studies, cross-modal maps are first acquired and then exploited to plan the behavior of robots. Morasso and Sanguineti [138] have proposed a model of body schema for motor planning that is presumably carried

---

[7]This work resembles the experiments by Nabeshima *et al.* [95] that we have encountered in Section III-C2. However, unlike Nabeshima, Hikita's work is a more direct attempt at modeling the putative biological mechanisms.
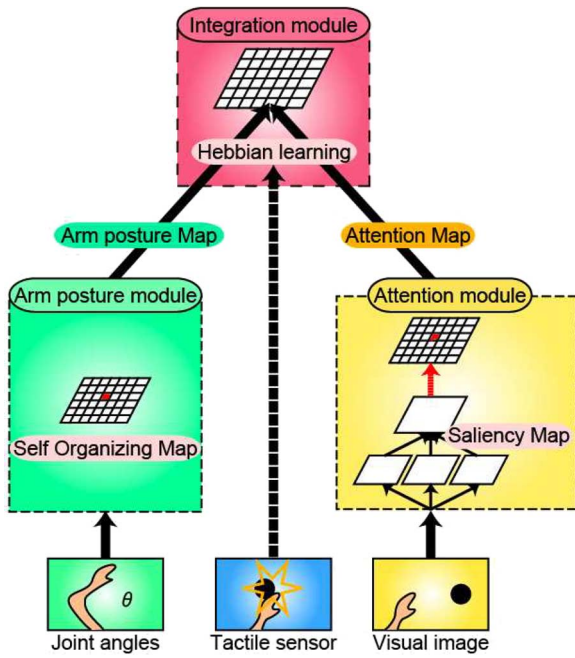
Fig. 7. Overview of the model proposed by Hikita *et al.* [96]. The association between the posture of the robot's arm and position in the visual field is triggered by tactile stimulation. A saliency map makes a robot fixate a point of contact between its end-effector and an object, since more salient features are observed at that point.
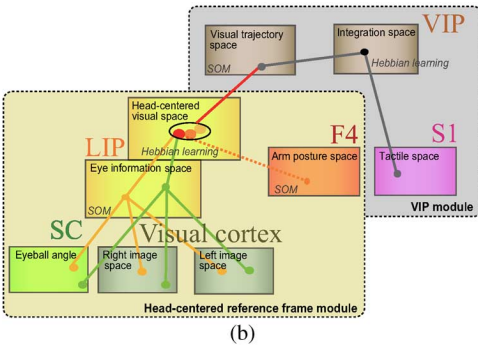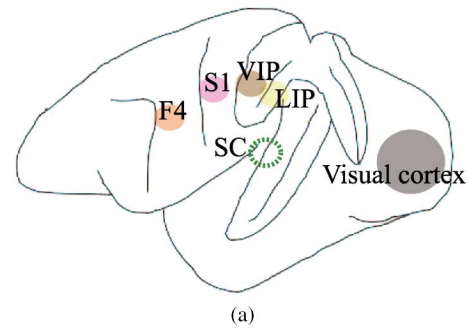

Fig. 9. Correspondence between brain regions (a) and representation spaces (b) proposed by Fuke *et al.* [134]. Eye information space is combined with arm posture space into a head-centered visual space. This process bears similarity to connections among F4 and LIP areas. Integration of the head-centered visual space and tactile space produce neural activities similar to the ones observed in VIP area.
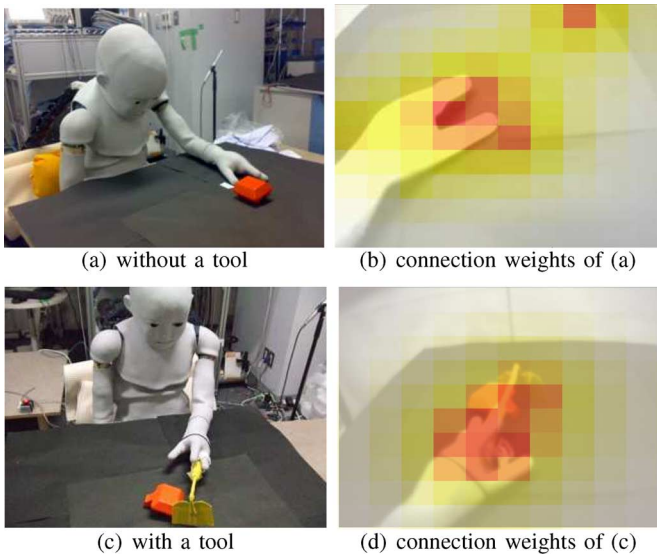

Fig. 8. Body schema extension during tool use. The connection between visual and proprioceptive fields: (a) and (b) without a tool; (c) and (d) with a tool (from Hikita *et al.* [96]). The red areas show a strong connection between visual and proprioceptive spaces in each setting. Confront Fig. 1 for the results from monkeys. (a) Without a tool; (b) connection weights of (a); (c) with a tool; (d) connection weights of (c).

out in area 5 of the posterior parietal cortex in association with the basal ganglia. The model is called SO-BoS (self-organizing body-schema) and consists of two components: a sensorimotor mapping (forward kinematic model implemented as a self-organizing cortical map), and inverse sensorimotor mapping (inverse kinematic model implemented as a gradient-descent mechanism in a potential field). The former is first acquired through motor

babbling. Then, the latter is tuned depending on the task constraints such as the target position and the posture to reach a target. Results of a simulation of a 3-degree-of-freedom arm show that the proposed model can realize different reaching behaviors that satisfy the constraints. However, the platform used is rather simplistic and the work has a more computational than synthetic modeling flavor.

Stoytchev [143] extended this model to a tool use scenario. An offset vector was added that represented the distance in the visual field from a position of an end-effector to the tip of a tool attached to the end-effector. Results with a simulation of a 2-degree-of-freedom arm showed that the proposed model can extend the body representation and successfully approach a visual target using the tool. The author has also shown that this model can acquire an extended body representation that allows the robot to guide its arm movements through a TV image. The robot detects its own body part based on the synchronization between its own movements and the changes of visual features on the TV screen [139].

Pitti *et al.* [140] have encoded body representations as a spatio–temporal pattern of neural activities in sensorimotor networks. Coordinated behavior, induced by morphological properties, was produced. Spiking neural networks were used to acquire mappings from a log-polar representation combined with a saliency map to motor commands for controlling the neck and the camera's orientation. Connections were regulated by spike-timing-dependent plasticity. Interaction among the body, environment, and the nervous system enabled a robot to self-organize the fixation behavior and the saccade behavior to a salient object. Analysis of the neural activities in the networks
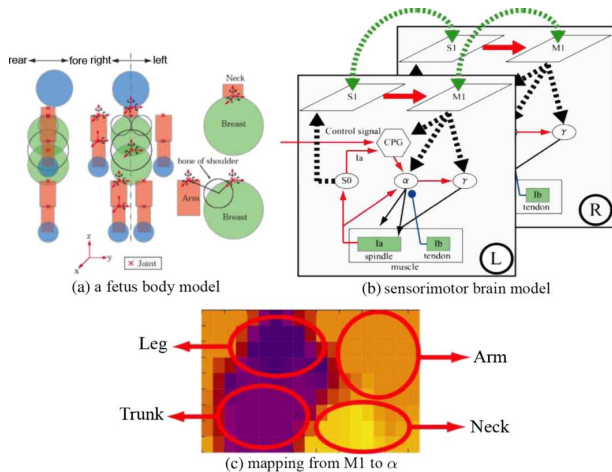
Fig. 10. Low-level body representations in a fetal model (Kuniyoshi and Sangawa [141]) (a) a fetus body model. Its physical properties such as size, mass and joint angle limitations were based on biological findings. (b) Cortico-medular-spinal-muscular model. (c) Self-organizing map from M1 to α, which displayed separation into areas corresponding to different body parts through spontaneous movement driven by the activities of the CPGs.

revealed a distinction between movement caused by the agent itself and that caused externally, thus representing a sense of agency (cf. Sections II-C1 and III-D2).

C. Development of a Low-Level Body Schema

Kuniyoshi and Sangawa [141] investigated the role of tight coupling between a body and its environment and how consistent dynamical patterns can emerge from this close physical interaction. They proposed a model of a neuro-musculo-skeletal system that consists of biologically realistic components such as a skeleton, muscles, spindles, tendon organs, spinal circuits, medullar circuits, and central pattern generators (CPGs). On top of that, a basic cortical model from self-organizing maps was constructed. The connections were modulated by Hebbian learning rule during spontaneous movement driven by the activities of the lower circuits. Self-organized body movement was observed in a simple musculo-skeletal model which consisted of two rigid objects connected with a free joint and multiple muscle fibers. This mediated the acquisition of low-level body representations, such as the relations between agonist and antagonist muscles. Further experiments with a human fetal model showed that simple movements, such as crawling and rolling, can emerge. The cortical maps displayed a separation into areas corresponding to different body parts shown in Fig. 10. Related to this work, a real robot that has anthropomimetic features is currently developed in the context of the ECCEROBOT project, where the development of a body schema will be subject to investigation [144].

V. Conclusion and Future Prospects

The research in cognitive sciences deals with many body representations that are short-term or long-term, conscious or unconscious, perception- or action-oriented. Synthesizing the putative mechanisms behind the biological body representations in robots can serve two goals. First, it can help to endow the robots with new capabilities that are ubiquitous in nature, yet unattainable by the machines of today. Second, synthetic modeling, i.e.,

investigating hypothetical mechanisms in artificial brain-body-environment systems, can complement empirical studies of psychology and neurosciences. These two avenues have been the subject of the present review.

To this end, we have first presented a review of the treatment of body representations in cognitive sciences. However, our survey was biased by having a robotic implementation in mind. Body representations in robots cover only a subspace of their biological counterparts so far. They can be long-term or short-term, but they can hardly be considered conscious and they are largely action-oriented (since we are usually interested in the robots performing some task). Therefore, we have largely neglected the phenomenological, or reflective, mechanisms of body representations, but concentrated on more low-level, pre-reflective, computational mechanisms such as plasticity of body representations, or coordinate transformations. We have explicitly attempted to clarify the relationship between the closely related notions of body schema, body image, peripersonal space and forward models.

To have a model of a robot in order to control it comes naturally to most control engineers and roboticists. A model of a plant (or robot) indeed is a representation that is used to guide the robot's actions and can thus be considered a kind of body schema. However, such a model has very different characteristics from those of a biological body schema: typically it is fixed, explicit, precise, centralized, and objective. These very characteristics of a classical model of a plant restrict the domains in which robots can be successfully used to very limited, precisely controlled environments. There are also costs associated with the development of such a model. Therefore, it is desirable that robots can develop, calibrate and adapt their models automatically. We have reviewed work that departs from the traditional field of robotics and extends it toward online automatic self-calibration. Beyond self-calibration, architectures that can also cope with topological changes have been analyzed, paving the way for the adaptive and the resilient machines of the future. Apart from body representations that have an explicit nature, the body schema of a robot can also be represented in an implicit manner. While this traditionally meant a connectionist (neural network) implementation, models using Bayesian networks are gaining popularity.

Both, explicit and implicit representations, have their pros and cons. Explicit models typically require more input from the designer—a parametrized kinematic model, or at least number and characteristics of joints, for instance. On the other hand, what they offer in return, can be possible integration with traditional control schemes, extrapolation to previously unseen configurations, or easier debugging. A representation that has an analytical form is also more compact and has an infinite resolution compared to a look-up table or self-organizing map that stores previously seen sensorimotor relationships only. The biggest merits of implicit representations probably are that little prior knowledge is required, and even problems that are outside of the scope of analytical treatment (e.g., deformable bodies) can be tackled. Calibration of sensorimotor mappings and their employment in control can be intertwined.

The mechanisms underlying the working and development of body schema (and body image) in animals and humans are still far from clear. Uncovering them has been largely the task of

neuroscience. Many findings were obtained by direct recordings from brain. However, even though the recording/imaging techniques are improving, there are still a lot of difficulties associated with "live" recordings from experimental subjects. Empirical studies have been supplemented by computational modeling. However, in many situations, a whole brain-body-environment system is indispensable. This is where robots and simulated robots come into play as tools to investigate biological body schema. While it is probably fair to say that many of the results are still preliminary, there are several relevant cases that we have reviewed.

We want to conclude by identifying the trends and also the weak spots in the research that we have just summarized and also propose areas for future research. First, the work on models in robotics is heavily biased toward manipulator arms, observed by a camera (cf. "humanoid torso" in Tables II and III). At the same time, the platforms are typically very stiff. This holds not only for traditional, but also for bio-inspired research. Therefore, a future research challenge is to deal with other behaviors and platforms: locomotion and compliant robots, for instance. Second, the integration of multiple modalities as demonstrated by biological agents, is still largely lacking—visual modality is often the only one that complements proprioception (joint angles). Third, next to traditional analytical methods from control theory and connectionist models, Bayesian networks are becoming a prominent tool to represent a body schema, with the additional benefits of integrating uncertainty in them. Fourth, most of the research discussed is demonstrated to work in rather simple scenarios (limited number of degrees of freedom, for instance). The extent to which the individual solutions can be scaled up is an open question.

REFERENCES

[1] H. Head and H. G. Holmes, "Sensory disturbances from cerebral lesions," *Brain*, vol. 34, pp. 102–254, 1911.

[2] F. de Vignemont, "Body schema and body image—Pros and cons," *Neuropsychologia*, vol. 48, no. 3, pp. 669–680, 2010.

[3] S. Gallagher, *How the Body Shapes the Mind*. London, U.K.: Oxford Univ. Press, 2005.

[4] H. Dijkerman and E. de Haan, "Somatosensory processes subserving perception and action," *Behav. Brain Sci.*, vol. 30, no. 2, pp. 189–201, 2007.

[5] J. Paillard, "Body schema and body image—A double dissociation in deafferented patients," Motor Control, Today and Tomorrow G. N. Gantchev, S. Mori, and J. Massion, Eds., 1999.

[6] D. Milner and M. Goodale, *The Visual Brain in Action*. London, U.K.: Oxford Univ. Press, 1995.

[7] L. G. Ungerleider and M. Mishkin, "Two cortical visual systems," in *Analysis of Visual Behavior*, D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield, Eds. Cambridge, MA: MIT Press, 1982, pp. 549–586.

[8] S. Creem and D. Proffitt, "Defining the cortical visual systems: "What," "where," and "how"," *Acta Psychologica*, vol. 107, no. 1–3, pp. 43–68, 2001.

[9] J. Schwoebel and H. B. Coslett, "Evidence for multiple, distinct representations of the human body," *J. Cogn. Neurosci.*, vol. 17, no. 4, pp. 543–553, 2005.

[10] P. Rochat, "Self-perception and action in infancy," *Exp. Brain Res.*, vol. 123, pp. 102–109, 1998.

[11] P. Rochat and S. J. Hespos, "Differential rooting response by neonates: Evidence for an early sense of self," *Early Develop. Parent.*, vol. 6, pp. 105–112, 1997.

[12] E. Bushnell and J. Boudreau, "Motor development and the mind: The potential role of motor abilities as a determinant of aspects of perceptual development," *Child Develop.*, vol. 64, no. 4, pp. 1005–1021, 1993.

[13] N. David, A. Newen, and K. Vogeley, "The "sense of agency" and its underlying cognitive and neural mechanisms," *Conscious. Cogn.*, vol. 17, pp. 523–534, 2008.

[14] M. Tsakiris, S. Schutz-Bosbach, and S. Gallagher, "On agency and body-ownership: Phenomenological and neurocognitive reflections," *Conscious. Cogn.*, vol. 16, pp. 645–660, 2007.

[15] G. Rizzolatti, C. Sinigaglia, and F. Anderson, *Mirrors in the Brain: How Our Minds Share Actions and Emotions*. London, U.K.: Oxford Univ. Press, 2007.

[16] V. S. Ramachandran and S. Blakeslee, *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. New York: William Mollow, 1998.

[17] N. Holmes and C. Spence, "Beyond the body schema: Visual, prosthetic, and technological contributions to bodily perception and awareness," in *Human Body Perception From the Inside Out*, G. Knoblich, I. Thornton, M. Grosjean, and M. Shiffrar, Eds. London, U.K.: Oxford Univ. Press, 2005, pp. 15–64.

[18] M. Botvinick and J. Cohen, "Rubber hands "feel" touch that eyes see," *Nature*, vol. 391, no. 6669, p. 756, Feb. 1998.

[19] M. Graziano, "Where is my arm? The relative role of vision and proprioception in the neuronal representation of limb position," *Proc. Nat. Acad. Sci.*, vol. 96, pp. 10418–10421, 1999.

[20] F. Pavani and U. Castiello, "Binding personal and extrapersonal space through body shadows," *Nat. Neurosci.* vol. 7, no. 1, pp. 14–16, Jan. 2004 [Online]. Available: http://dx.doi.org/10.1038/nn1167

[21] K. Carrie Armel and V. S. Ramachandran, "Projecting sensations to external objects: Evidence from skin conductance response," *Proc. Biol. Sci.*, vol. 270, no. 1523, pp. 1499–1506, Jul. 2003.

[22] H. H. Ehrsson, C. Spence, and R. E. Passingham, "That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb," *Science*, vol. 305, no. 5685, pp. 875–877, Aug. 2004.

[23] H. Ehrsson, "Rubber hand illusion," in *Oxford Companion to Consciousness*, T. Bayne, A. Cleermans, and P. Wilken, Eds. London, U.K.: Oxford Univ. Press, 2009, pp. 531–573.

[24] H. H. Ehrsson, B. Rosen, A. Stockselius, C. Ragno, P. Köhler, and G. Lundborg, "Upper limb amputees can be induced to experience a rubber hand as their own," *Brain*, vol. 131, pt. 12, pp. 3443–3452, Dec. 2008.

[25] H. H. Ehrsson, K. Wiech, N. Weiskopf, R. J. Dolan, and R. E. Passingham, "Threatening a rubber hand that you feel is yours elicits a cortical anxiety response," *Proc. Nat. Acad. Sci.*, vol. 104, no. 23, pp. 9828–9833, Jun. 2007.

[26] K. Hägni, K. Eng, M. Hepp-Reymond, L. Holper, B. Keisker, E. Siekierka, and D. Kiper, "Observing virtual arms that you imagine are yours increases the galvanic skin response to an unexpected threat," *PLoS ONE*, vol. 3, no. 8, p. e3082, 08 2008.

[27] A. Iriki, M. Tanaka, and Y. Iwamura, "Coding of modified body schema during tool use by macaque postcentral neurones," *Cogn. Neurosci. Neuropsychol.*, vol. 7, no. 14, pp. 2325–2330, 1996.

[28] N. P. Holmes and C. Spence, "The body schema and the multisensory representation(s) of peripersonal space," *Cogn. Process.*, vol. 5, no. 2, pp. 94–105, Jun. 2004.

[29] A. Maravita, C. Spence, and J. Driver, "Multisensory integration and the body schema: Close to hand and within reach," *Current Biol.*, vol. 13, no. 13, pp. R531–R539, Jul. 2003.

[30] S. H. Johnson-Frey, "What's so special about human tool use?," *Neuron*, vol. 39, pp. 201–204, 2003.

[31] A. Maravita and A. Iriki, "Tools for the body (schema)," *Trends Cogn. Sci.*, vol. 8, no. 2, pp. 79–86, Feb. 2004.

[32] A. Iriki, "Posterior parietal cortex and tool usage and hand shape," *Encyclopedia Neurosci.*, vol. 7, pp. 797–802, 2009.

[33] A. Berti and F. Frassinetti, "When far becomes near: Remapping of space by tool use," *J. Cogn. Neurosc.*, vol. 12, no. 3, pp. 415–420, 2000.

[34] N. Holmes, G. Calvert, and C. Spence, "Extending or projecting peripersonal space with tools? Multisensory interactions highlight only the distal and proximal ends of tools," *Neurosci. Lett.*, vol. 372, pp. 62–67, 2004.

[35] M. Slater, D. Perez-Marcos, H. H. Ehrsson, and M. V. Sanchez-Vives, "Inducing illusory ownership of a virtual body," *Frontiers Neurosci.*, vol. 3, pp. 214–220, 2009.

[36] J. Milla'n and D. R. , "Adaptive brain interfaces," *Commun. ACM*, vol. 46, pp. 74–80, 2003.

[37] M. A. Lebedev and M. A. L. Nicolelis, "Brain—Machine interfaces: Past, present and future," *Trends Neurosci.*, vol. 29, no. 9, pp. 536–546, 2006.

[38] D. J. McFarland and J. R. Wolpaw, "Brain-computer interface operation of robotic and prosthetic devices," *Computer*, vol. 41, pp. 52–56, Oct. 2008.

[39] J. M. Carmena, M. A. Lebedev, R. E. Crist, J. E. O'Doherty, D. M. Santucci, D. F. Dimitrov, P. G. Patil, C. S. Henriquez, and M. A. L. Nicolelis, "Learning to control a brain-machine interface for reaching and grasping by primates," *PLoS Biol.*, vol. 1, no. 2, p. E42, Nov. 2003.

[40] M. Velliste, S. Perel, M. C. Spalding, A. S. Whitford, and A. B. Schwartz, "Cortical control of a prosthetic arm for self-feeding," *Nature*, vol. 453, no. 7198, pp. 1098–1101, 2008.

[41] A. Hernandez Arieta, R. Kato, H. Yokoi, T. Arai, and T. Ohnishi, "An fmri study on the effects of electrical stimulation as biofeedback," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, Beijing, China, Oct. 2006, pp. 4336–4342.

[42] A. Hernandez Arieta, K. Dermitzakis, D. Damian, M. Lungarella, and R. Pfeifer, "Sensory-motor coupling in rehabilitation robotics," in *Handbook of Service Robotics*. Rijeka, Croatia: I-Tech Education, 2008, pp. 21–36.

[43] C. Cipriani, C. Antfolk, C. Balkenius, B. Rosén, G. Lundborg, M. C. Carrozza, and F. Sebelius, "A novel concept for a prosthetic hand with a bidirectional interface: A feasibility study," *IEEE Trans. Biomed. Eng.* vol. 56, no. 11, pt. 2, pp. 2739–2743, Nov. 2009 [Online]. Available: http://dx.doi.org/10.1109/TBME.2009.2031242

[44] J. C. Sanchez, B. Mahmoudi, J. DiGiovanna, and J. C. Principe, "Exploiting co-adaptation for the design of symbiotic neuroprosthetic assistants," *Neural Netw.*, vol. 22, pp. 305–315, 2009.

[45] J. DiGiovanna, B. Mahmoudi, J. Fortes, J. C. Principe, and J. C. Sanchez, "Coadaptive brain-machine interface via reinforcement learning," *IEEE Trans. Biomed. Eng.* vol. 56, no. 1, pp. 54–64, Jan. 2009 [Online]. Available: http://dx.doi.org/10.1109/TBME.2008.926699

[46] R. A. Andersen, R. M. Bracewell, S. Barash, J. W. Gnadt, and L. Fogassi, "Eye position effects on visual memory, and saccade-related activity in areas lip and 7a of macaque," *J. Neurosci.*, vol. 10, no. 4, pp. 1176–1196, 1990.

[47] J. R. Duhamel, C. L. Colby, and M. E. Goldberg, "Ventral intraparietal area of the macaque: Congruent visual and somatic response properties," *J. Neurophysiol.*, vol. 79, pp. 126–136, 1998.

[48] M. I. Sereno and R. Huang, "A human parietal face area contains aligned head-centered visual and tactile maps," *Nature Neurosci.*, vol. 9, pp. 1337–1343, 2006.

[49] G. J. Blatt, R. A. Andersen, and G. R. Stoner, "Visual receptive field organization and cortico-cortical connections of the lateral intraparietal area (area lip) in the macaque," *J. Compar. Neurol.*, vol. 299, no. 4, pp. 421–445, 1990.

[50] J.-R. Duhamel, F. Bremmer, S. BenHamed, and W. Graf, "Spatial invariance of visual receptive fields in parietal cortex neurons," *Nature*, vol. 389, pp. 845–848, 1997.

[51] A. P. Batista, C. A. Buneo, L. H. Snyder, and R. A. Andersen, "Reach plans in eye centred coordinates," *Science*, vol. 285, pp. 257–260, 1999.

[52] O. Mullette-Gillman, Y. E. Cohen, and J. M. Groh, "Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus," *J. Neurophysiol.*, vol. 94, no. 4, p. 2331, 2005.

[53] S. W. Chang, C. Papadimitriou, and L. H. Snyder, "Using a compound gain field to compute a reach plan," *Neuron*, vol. 64, pp. 744–755, 2009.

[54] C. A. Buneo, M. R. Javis, A. P. Batista, and R. A. Andersen, "Direct visuomotor transformations for reaching," *Nature*, vol. 416, pp. 632–636, 2002.

[55] G. Blohm and J. D. Crawford, "Fields of gain in the brain," *Neuron*, vol. 64, pp. 598–600, 2009.

[56] E. Salinas and L. F. Abbot, "Coordinate transformations in the visual system: How to generate gain fields and what to compute with them," *Progress Brain Res.*, vol. 130, pp. 175–190, 2001.

[57] A. Pouget, S. Deneve, and J.-R. Duhamel, "A computational perspective on the neural basis of multisensory spatial representations," *Nature Rev. Neurosci.*, vol. 3, pp. 741–747, 2002.

[58] L. H. Snyder, K. L. Grieve, P. Brotchie, and R. A. Andersen, "Separate body- and world-referenced representations of visual space in parietal cortex," *Nature*, vol. 394, pp. 887–891, 1998.

[59] M. Kawato, "Internal models for motor control and trajectory planning," *Current Opinion Neurobiol.*, vol. 9, pp. 718–727, 1999.

[60] P. Davidson and D. M. Wolpert, "Widespread access to predictive models in the motor system: A short review," *J. Neural Eng.*, vol. 2, pp. 313–319, 2005.

[61] R. Grush, "The emulation theory of representation—Motor control, imagery, and perception," *Behav. Brain Sci.*, vol. 27, pp. 377–442, 2004.

[62] H. Marques and O. Holland, "Architectures for functional imagination," *Neurocomputing*, vol. 72, pp. 743–759, 2009, in press.

[63] J. K. O'Regan and A. Noe, "A sensorimotor account of vision and visual consciousness," *Behav. Brain Sci.*, vol. 24, pp. 939–1031, 2001.

[64] Z. Pylyshyn, *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, Z. Pylyshyn, Ed. New York: Ablex, 1987.

[65] R. A. Brooks, "Intelligence without representation," *Artif. Intell. J.*, vol. 47, pp. 139–159, 1991.

[66] R. Pfeifer and C. Scheier, *Understanding Intelligence*. Cambridge, MA: MIT Press, 1999.

[67] R. Pfeifer, M. Lungarella, and F. Iida, "Self-organization, embodiment, and biologically inspired robotics," *Science*, vol. 318, pp. 1088–1093, 2007.

[68] G. W. Walter, *The Living Brain*, G. W. Walter, Ed. New York: Norton, 1953.

[69] V. Braitenberg, *Vehicles—Experiments in Synthetic Psychology*, V. Braitenberg, Ed. Cambridge, MA: MIT Press, 1986.

[70] R. Brooks, "A robust layered control system for a mobile robot," *IEEE J. Robotic. Autom.*, vol. 2, no. 1, pp. 14–23, Jan. 1986.

[71] R. A. Brooks, "A robot that walks: Emergent behaviors from a carefully evolved network," *Neural Comput.*, vol. 1, pp. 153–162, 1989.

[72] H. Cruse, T. Kindermann, M. Schumm, J. Dean, and J. Schmitz, "Walknet—A biologically inspired network to control six-legged walking," *Neural Netw.*, vol. 11, pp. 1435–1447, 1998.

[73] L. Sciavicco, B. Siciliano, and B. Sciavicco, *Modelling and Control of Robot Manipulators*, L. Sciavicco, B. Siciliano, and B. Sciavicco, Eds. Secaucus, NJ: Springer-Verlag, 2000.

[74] M. Desmurget and S. Grafton, "Forward modeling allows feedback control for fast reaching movements," *Trends Cogn. Sci.*, vol. 4, no. 11, pp. 423–431, 2000.

[75] B. Webb, "Neural mechanisms for prediction: Do insects have forward models?," *Trends Neurosci.*, vol. 27, pp. 278–282, 2004.

[76] J. Heikkonen and P. Koikkalainen, "Self-organization and autonomous robots," in *Neural Systems for Robotics*. New York: Academic, 1997, pp. 297–337.

[77] L. L. E. Massone, "Sensorimotor learning," in *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA: MIT Press, 1995, pp. 860–864.

[78] D. M. Wolpert, R. C. Miall, and M. Kawato, "Internal models in the cerebellum," *Trends Cogn. Sci.*, vol. 2, no. 9, pp. 338–347, 1998.

[79] F. L. Lewis, C. T. Abdallah, and D. M. Dawson, *Control of Robot Manipulators*. New York: Macmillian, 1993.

[80] J. Craig, *Introduction to Robotics: Mechanics and Control*, J. Craig, Ed. Boston, MA: Addison-Wesley Longman Publishing, 1989.

[81] M. Spong and M. Vidyasagar, *Robot Dynamics and Control*, M. Spong and M. Vidyasagar, Eds. New York: Wiley, 1989.

[82] M. Hersch, E. Sauser, and A. Billard, "Online learning of the body schema," *Int. J. Humanoid Robot.*, vol. 5, pp. 161–181, 2008.

[83] J. Sturm, C. Plagemann, and W. Burgard, "Body schema learning for robotic manipulators from visual self-perception," *J. Physiol.—Paris*, vol. 103, pp. 220–231, 2009.

[84] E. Alpaydin, *Introduction to Machine Learning*, E. Alpaydin, Ed. Cambridge, MA: MIT Press, 2004.

[85] K. Kozlowski, *Modelling and Identification in Robotics*, K. Kozlowski, Ed. Berlin, Germany: Springer-Verlag, 1998.

[86] L. Ljung, *System Identification: Theory for the User*, L. Ljung, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1999.

[87] N. Roy and S. Thrun, "Online self-calibration for mobile robots," in *Proc. of the IEEE Int. Conf. Robot. Autom. (ICRA)*, Detroit, MI, 1999.

[88] D. Bennett, D. Geiger, and J. Hollerbach, "Autonomous robot calibration for hand-eye coordination," *Int. J. Robot. Res.*, vol. 10, no. 5, pp. 550–559, 1991.

[89] J. Hollerbach and C. Wampler, "The calibration index and taxonomy for robotic kinematic calibration methods," *Int. J. Robot. Res.*, vol. 15, no. 6, pp. 573–591, 1996.

[90] C. Gatla, R. Lumia, J. Wood, and G. Starr, "An automated method to calibrate industrial robots using a virtual closed kinematic chain," *IEEE Trans. Robot.*, vol. 23, no. 6, pp. 1105–1116, Dec. 2007.

[91] R. Tsai and R. Lenz, "Real time versatile robotics hand/eye calibration using 3d machine vision," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, 1988.

[92] J. Bongard and H. Lipson, "Nonlinear system identification using co-evolution of models and tests," *IEEE Trans. Evol. Comput.*, vol. 9, no. 4, pp. 361–384, Aug. 2005.

[93] M. Hersch, "Adaptive Sensorimotor Peripersonal Space Representation and Motor Learning for a Humanoid Robot," Ph.D. dissertation, EPFL Lausanne, Lausanne, Switzerland, 2009.

[94] R. Martinez-Cantin, M. Lopes, and L. Montesano, "Body schema acquisition through active learning," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, Anchorage, AK, 2010.

[95] C. Nabeshima, Y. Kuniyoshi, and M. Lungarella, "Adaptive body schema for robotic tool-use," *Adv. Robot.*, vol. 20, no. 11, pp. 1105–1126, 2006.

[96] M. Hikita, S. Fuke, M. Ogino, T. Minato, and M. Asada, "Visual attention by saliency leads cross-modal body representation," in *Proc. 7th Int. Conf. Develop. Learn. (ICDL)*, Shanghai, China, 2009.

[97] J. Bongard, V. Zykov, and H. Lipson, "Resilient machines through continuous self-modeling," *Science*, vol. 314, pp. 1118–1121, 2006.

[98] J. Bongard, V. Zykov, and H. Lipson, "Automated synthesis of body schema using multiple sensor modalities," in *Proc. Int. Conf. Simul. Synth. Living Syst. (ALIFEX)*, Bloomington, IN, 2006.

[99] G. Metta, G. Sandini, and J. Konczak, "A developmental approach to visually-guided reaching in artificial systems," *Neural Netw.*, vol. 12, no. 10, pp. 1413–1427, 1999.

[100] C. Gaskett and G. Cheng, "Online learning of a motor map for humanoid robot reaching," in *Proc. 2nd Int. Conf. Computat. Intell., Robot. Autonom. Syst. (CIRAS 2003)*, Singapore, 2003.

[101] Y. Yoshikawa, K. Hosoda, and M. Asada, "Does the invariance in multi-modalities represent the body scheme?—A case study with vision and proprioception—," in *Proc. 2nd Int. Symp. Adapt. Motion Animals Mach. Volume SaP-II-1*, Kyoto , Japan, 2003.

[102] K. Gold and B. Scassellati, "Using probabilistic reasoning over time to self-recognize," *Robot. Autonom. Syst.*, vol. 57, no. 4, pp. 384–392, 2009.

[103] L. Natale, F. Orabona, G. Metta, and G. Sandini, "Sensorimotor coordination in a "baby" robot: Learning about objects through grasping," *Progress Brain Res.*, vol. 164, pp. 403–424, 2007.

[104] A. Dearden and Y. Demiris, "Learning forward models for robots," in *Proc. Int. Conf. Artif. Intell. (IJCAI)*, Edinburgh, Scotland, 2005.

[105] D. Grimes, R. Chalodhorn, and R. Rao, "Dynamic imitation in a humanoid robot through nonparametric probabilistic inference," in *Proc. Robot.: Sci. Syst. (RSS)*, Philadelphia, PA, 2006.

[106] M. I. Jordan and D. E. Rumelhart, "Forward models: Supervised learning with a distal teacher," *Cogn. Sci.*, vol. 16, pp. 307–354, 1992.

[107] A. D'Souza, S. Vijayakumar, and S. Schaal, "Learning inverse kinematics," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Lausanne, Switzerland, 2001.

[108] S. M. Prabhu and D. P. Garg, "Artificial neural network based robot control: An overview," *J. Intell. Robot. Syst.*, vol. 15, pp. 333–365, 1996.

[109] Q. Meng and M. Lee, "Automated cross-modal mapping in robotic eye/hand systems using plastic radial basis function networks," *Connect. Sci.*, vol. 19, no. 1, pp. 25–52, 2007.

[110] M. Hagan and H. Demuth, "Neural networks for control," in *Proc. Amer. Contr. Conf.*, San Diego, CA, 1999.

[111] J. S. Albus, "A new approach to manipulator control: The cerebellar model articulation controller (cmac)," *J. Dynamic Syst., Measure., Contr.*, vol. 97, pp. 220–227, 1975.

[112] Z. Jiang and S. Wang, "A general learning scheme for cmac-based controller," *Neural Process. Lett.*, vol. 18, no. 2, pp. 125–138, 2003.

[113] W. Miller, R. Sutton, and P. Werbos, *Neural Networks for Control*. Cambridge, MA: MIT Press, 1990.

[114] K. J. Hunt, D. Sbarbaro, R. Zbikovski, and P. J. Gawthrop, "Neural networks for control systems—A survey," *Automatica*, vol. 28, no. 6, pp. 1083–1112, 1992.

[115] G. d. A. Barreto, A. F. R. Araujo, and H. Ritter, "Self-organizing feature maps for modeling and control of robotic manipulators," *J. Intell. Robot. Syst.*, vol. 36, pp. 407–450, 2003.

[116] M. Kawato, K. Furukawa, and R. Suzuki, "A hierarchical neural-network model for control and learning of voluntary movement," *Biol. Cybern.*, vol. 57, pp. 169–185, 1987.

[117] M. Zeller, R. Sharma, and K. Schulten, "Motion planning of a pneumatic robot using a neural network," *IEEE Control Syst. Mag.*, vol. 17, pp. 89–98, 1997.

[118] L. Natale, F. Nori, G. Metta, and G. Sandini, "Learning precise 3d reaching in a humanoid robot," in *Proc. Int. Conf. Develop. Learn. (ICDL)*, London, U.K., 2007.

[119] S. Hutchinson, G. D. Hager, and P. I. Corke, "A tutorial on visual servo control," *IEEE Trans. Robot. Autom.*, vol. 12, no. 5, pp. 651–670, Oct. 1996.

[120] D. Mansard, M. Lopes, J. Santos-Victor, and F. Chaumette, "Jacobian learning methods for tasks sequencing in visual servoing," in *Proc. Int. Conf. Intell. Robot. Syst. (IROS)*, Beijing, China, 2006.

[121] M. Jones and D. Vernon, "Using neural networks to learn hand-eye co-ordination," *Neural Comput. Appl.*, vol. 2, pp. 2–12, 1994.

[122] M. Kuperstein, "Infant neural controller for adaptive sensory-motor coordination," *Neural Netw.*, vol. 4, pp. 131–145, 1991.

[123] S. Rougeaux and Y. Kuniyoshi, "Robust tracking by a humanoid vision system," in *Proc. 1st Int. Workshop Humanoid Human Friendly Robot.*, Tsukuba, Japan, 1998.

[124] Y. Yoshikawa, Y. Tsuji, K. Hosoda, and M. Asada, "Is it my body?—Body extraction from uninterpreted sensory data based on the invariance of multiple sensory attributes—," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, Sendai, Japan, 2004.

[125] P. Fitzpatrick and G. Metta, "Toward manipulation-driven vision," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Lausanne, Switzerland, 2002.

[126] C. Kemp and A. Edsinger, "What can i control? The development of visual categories for a robot's body and the world that it influences," in *Proc. 5th Int. Conf. Develop. Learn. (ICDL)*, New Delhi, India, 2006.

[127] M. Vukobratovic and B. Vorovac, "Zero-moment point—Thirty five years of its life," *Int. J. Humanoid Robot.*, vol. 1, no. 1, pp. 157–173, 2004.

[128] K. Hirai, M. Hirose, Y. Haikawa, and T. Takenaka, "The development of honda humanoid robot," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Leuven, Beligum, 1998.

[129] J. Yamaguchi, E. Soga, S. Inoue, and A. Takanishi, "Development of bipedal humanoid robot-control method of wholebody cooperative dynamic biped walking," in *Proc. IEEE Int. Conf. Robot. Automation (ICRA)*, Detroit, MI, 1999.

[130] K. Nishiwaki, J. J. Kuffner, S. Kagami, M. Inaba, and H. Inoue, "The experimental humanoid robot h7: A research platform for autonomous behavior," *Phil. Trans. Roy. Soc. A*, vol. 365, no. 1850, pp. 79–107, 2007.

[131] J. Buchli, M. Kalakrishnan, M. Mistry, P. Pastor, and S. Schaal, "Compliant quadruped locomotion over rough terrain," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, St. Louis, MO, 2009, pp. 814–820.

[132] R. Pfeifer and C. Scheier, *Understanding Intelligence*. Cambridge, MA: MIT Press, 2001.

[133] M. Asada, K. Hosoda, Y. Kuniyoshi, H. Ishiguro, T. Inui, Y. Yoshikawa, M. Ogino, and C. Yoshida, "Cognitive developmental robotics: A survey," *IEEE Trans. Autonom. Mental Develop.*, vol. 1, no. 1, pp. 12–34, May 2009.

[134] S. Fuke, M. Ogino, and M. Asada, "Acquisition of the head-centered peri-personal spatial representation found in vip neuron," *IEEE Trans. Autonom. Mental Develop.*, vol. 1, no. 2, pp. 110–140, Aug. 2009.

[135] S. Fuke, M. Ogino, and M. Asada, "Body image constructed from motor and tactile images with visual information," *Int. J. Human. Robot.*, vol. 4, pp. 347–364, 2007.

[136] Y. Yoshikawa, H. Kawanishi, M. Asada, and K. Hosoda, "Body scheme acquisition by cross modal map learning among tactile, visual, and proprioceptive spaces," in *Proc. 2nd Int. Workshop Epigen. Robot.: Model. Cogn. Develop. Robot. Syst.*, Edinburgh, Scotland, 2002, pp. 181–184.

[137] A. Pitti, H. Alirezaei, and Y. Kuniyoshi, "Cross-modal and scale-free action representations through enaction," *Neural Netw.*, vol. 22, no. 2, pp. 144–154, 2009.

[138] P. Morasso and V. Sanguineti, "Self-organizing body schema for motor planning," *J. Motor Behav.*, vol. 27, no. 1, pp. 52–66, 1995.

[139] A. Stoytchev, "Toward video-guided robot behaviors," in *Proc. 7th Int. Conf. Epigen. Robot.*, Piscataway, NJ, 2007, pp. 165–172.

[140] A. Pitti, H. Mori, S. Kouzuma, and Y. Kuniyoshi, "Contingency perception and agency measure in visuo-motor spiking neural networks," *IEEE Trans. Autonom. Mental Develop.*, vol. 1, no. 1, pp. 86–97, May 2009.

[141] Y. Kuniyoshi and S. Sangawa, "Early motor development from partially ordered neural-body dynamics: Experiments with a. cortico-spinal-musculo-sleletal model," *Biol. Cybern.*, vol. 95, pp. 589–605, 2006.

[142] L. Itti and F. Pighin, "Realistic avatar eye and head animation using a neurobiological model of visual attention," in *Proc. SPIE 48th Annu. Int. Symp. Optical Sci. Technol.*, Denver, CO, 2004, vol. 5200, pp. 64–78.

[143] *Computational Model for an Extendable Robot Body Schema*, GIT-CC-03-44, College of Computing, Georgia Institute of Technology, 2003.

[144] H. Marques, M. Jantsch, S. Wittmeier, O. Holland, C. Alessandro, M. Lungarella, and R. Knight, "Ecce1: The first of a series of anthropomimetic musculoskelal upper torsos," in *Proc. Int. Conf. Humanoids*, Nashville, TN, 2010.

Matej Hoffmann received the M.Sc. degree in computer science from Charles University, Prague, Czech Republic, in 2006. He is currently working towards the Ph.D. degree in artificial intelligence at the Artificial Intelligence Laboratory, University of Zurich, Zurich, Switzerland.

His research interests include legged locomotion, motor control, and embodied cognition in animals and robots.

Hugo Gravato Marques received the M.Sc. degree in informatics and computer engineering from the University of Porto, Porto, Portugal, in 2003. He received the Ph.D. degree from the University of Essex, Essex, U.K., in 2009.

He is currently a Postdoctoral Researcher at the AI Lab in the University of Zurich, Zurich, Switzerland, where he is investigating the control of a compliant humanoid upper torso. He is particularly interested in the field of developmental robotics and connecting his research to the spinal development in humans.

Alejandro Hernandez Arieta received the B.Sc. degree in electronics systems engineering from the Monterrey Institute of Technology and Higher Education, Queretaro, Mexico, in 1998, and the M.Sc. degree in systems and information engineering from Hokkaido University, Sapporo, Japan, in 2004. He also received the Ph.D. degree in precision engineering from the University of Tokyo, Tokyo, Japan, in 2007.

He is currently a Research Fellow at the Artificial Intelligence Laboratory of the University of Zurich, Zurich, Switzerland. From 2005 to 2007, he worked as a research assistant at the Artificial Intelligence Laboratory. He has published 26 scientific publications and has participated as an invited speaker at several colloquiums and seminars. His research interests include robot technology, assistive devices, adaptive learning, functional electrical stimulation, rehabilitation and prosthetic devices.

Dr. Hernandez Arieta is a member of the IEEE Society of Engineering in Medicine and Biology.
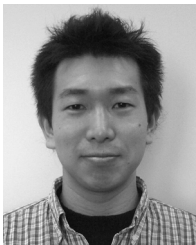
Hidenobu Sumioka received the B.Eng., M.Eng., and Ph.D. degrees in engineering from Osaka University, Osaka, Japan, in 2004, 2005, and 2008, respectively.

From April 2008 to March 2010, he was a Research Fellow of the Japan Society for the Promotion of Science (JSPS fellow). Since April 2010, he has been a Researcher at the Artificial Intelligence Laboratory, University of Zurich, Zurich, Switzerland. His research interests include emergence of behavior, human–robot interaction, joint attention, contingency detection, and cognitive developmental robotics.

Max Lungarella received the M.Eng. degree in electrical engineering from the University of Perugia, Perugia, Italy, in 1999, and the Ph.D. degree from the University of Zurich, Zurich, Switzerland, in 2004.

He is currently a Senior Researcher at the University of Zurich. He is also Chief Technology Officer of Dynamic Devices LLC. From 2002 to 2004, he was an Invited Researcher at the Neuroscience Research Institute of the National Institute of Advanced Industrial Science and Technology in Tsukuba, Japan. From 2004 to 2007, he worked at the Department of Mechano-Informatics of the University of Tokyo, first as a Research Associate, then as a JSPS Postdoctoral Researcher, and as an ERAVO research fellow. He has been involved in various projects related to intelligent robotic systems and has organized a number of conferences such as "The 50th Anniversary World Summit of Artificial Intelligence" in 2006. His research interests include artificial intelligence, artificial life, network and information theory, computational biology, design automation, and robot technology.

Rolf Pfeifer received the M.Sc. degree in physics and mathematics, and the Ph.D. degree in computer science from the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, in 1970 and 1979, respectively.

Since 1987, he has been a Professor of Computer Science at the Department of Informatics, University of Zurich, Zurich, Switzerland, and Director of the Artificial Intelligence Laboratory. Having worked as a visiting Professor and Research Fellow at the Free University of Brussels, the MIT Artificial Intelligence Laboratory in Cambridge, MA, the Neurosciences Institute (NSI) in San Diego, CA, the Beijing Open Laboratory for Cognitive Science, and the Sony Computer Science Laboratory in Paris, he was elected "21st Century COE Professor, Information Science and Technology" at the University of Tokyo, Tokyo, Japan. In 2009, he was also a Visiting Professor at the Scuola Superiore Sant'Anna in Pisa, at Shanghai Jiao Tong University in China, and he was appointed "Fellow of the School of Engineering" at the University of Tokyo. He is the author of the book *Understanding Intelligence*, 1999 (with C. Scheier) and *How the Body Shapes the Way We Think: A New View of Intelligence*, 2007. His research interests are in the areas of embodiment, biorobotics, artificial evolution and morphogenesis, modular robotics, self-assembly, and educational technology.