RE-ENACTING SENSORIMOTOR EXPERIENCE FOR COGNITION

EDITED BY : Guido Schillaci, Verena V. Hafner and Bruno Lara

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: **researchtopics@frontiersin.org**
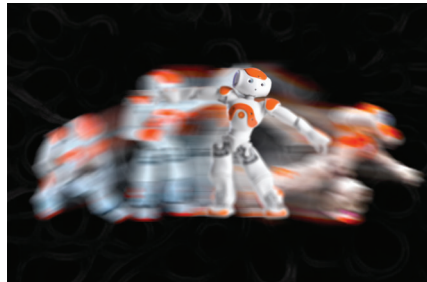
RE-ENACTING SENSORIMOTOR EXPERIENCE FOR COGNITION

Topic Editors:
Guido Schillaci, Humboldt-Universität zu Berlin, Germany
Verena V. Hafner, Humboldt-Universität zu Berlin, Germany
Bruno Lara, Universidad Autónoma del Estado de Morelos, Mexico

Artificial agent, re-enacting past and future sensorimotor states.

NAO robot created by SoftBank Robotics

Mastering the sensorimotor capabilities of our body is a skill that we acquire and refine over time, starting at the prenatal stages of development. This learning process is linked to brain development and is shaped by the rich set of multimodal information experienced while exploring and interacting with the environment.

Evidence coming from neuroscience suggests the brain forms and maintains body representations as the main strategy to this mastering. Although it is still not clear how this knowledge is represented in our brain, it is reasonable to think that such internal models of the body undergo a continuous process of adaptation. They need to match growing corporal dimensions during development, as well as temporary changes in the characteristics of the body, such as the transient morphological alterations produced by the usage of tools.

In the robotics community there is an increasing interest in reproducing similar mechanisms in artificial agents, mainly motivated by the aim of producing autonomous adaptive systems that can deal with complexity and uncertainty in human environments. Although promising results have been achieved in the context of sensorimotor learning and autonomous generation of body representations, it is still not clear how such low-level representations can be scaled up to more complex motor skills and how they can enable the development of cognitive capabilities.

Recent findings from behavioural and brain studies suggests that processes of mental simulations of action-perception loops are likely to be executed in our brain and are dependent on internal motor representations. The capability to simulate sensorimotor experience might represent a key mechanism behind the implementation of further cognitive skills, such as self-detection, self-other distinction and imitation. Empirical investigation on the functioning of similar processes in the brain and on their implementation in artificial agents is fragmented.

This e-book comprises a collection of manuscripts published by Frontiers in Robotics and Artificial Intelligence, under the section Humanoid Robotics, on the research topic re-enactment of sensorimotor experience for cognition in artificial agents. This compendium aims at condensing the latest theoretical, review and experimental studies that address new paradigms for learning and integrating multimodal sensorimotor information in artificial agents, re-use of the sensorimotor experience for cognitive development and further construction of more complex strategies and behaviours using these concepts.

The authors would like to thank M.A. Dylan Andrade for his art work for the cover.

Table of Contents

Editorial: Re-Enacting Sensorimotor Experience for Cognition

Guido Schillaci[1]*, Verena V. Hafner[1] and Bruno Lara[2]

[1] *Adaptive Systems Group, Department of Computer Science, Humboldt-Universität zu Berlin, Berlin, Germany,* [2] *Cognitive Robotics Group, Center for Science Research, Universidad Autonoma del Estado de Morelos, Cuernavaca, Mexico*

Editorial on the Research Topic

Re-Enacting Sensorimotor Experience for Cognition

Recent findings in cognitive science suggest that the human brain implements processes of simulation of sensorimotor activity (Pezzulo et al., 2013; Case et al., 2015; Wood et al., 2016). By re-enacting sensorimotor experience, the brain would be capable of anticipating the sensory consequences of intended motor actions. This would enable the individual to efficiently and fluidly interact with the environment.

This e-book puts forward the hypothesis that similar mechanisms underlie the development of basic cognitive capabilities. Therefore, sensorimotor simulation processes may represent one of the bridges between motor development and cognitive development in humans.

This collection comprises manuscripts published by Frontiers in *Robotics and Artificial Intelligence*, under the section Humanoid Robotics in the research topic "Re-enacting sensorimotor experience for cognition." The e-book aims at condensing the latest theoretical review and experimental studies that address new paradigms for learning and integrating multimodal sensorimotor information in artificial agents, re-use of the sensorimotor experience for cognitive development and further construction of more complex strategies and behaviors using these concepts.

1. THEORETICAL AND REVIEW STUDIES

In their review paper, Schillaci et al. introduce recent research on exploration as a drive for motor and cognitive development, and how this has been applied to robotics. After focusing on the development of internal body representations, the authors review research that highlights the importance of sensorimotor simulations and their role in the grounding of higher cognitive capabilities in robots. Most of these works have been inspired by sensorimotor and enactive theories. Froese and Sierra, in their review of the volume edited by Bishop and Martin on Contemporary Sensorimotor Theory (2014), draw the attention of the reader to the similarities and the differences of the current sensorimotor and enactive theories. However, the authors point out the need of additional comparative studies, in particular in the context of Robotics and AI.

Nonetheless, several challenges have already been posed by these theories. How can we explain the phenomenological character of experience (Froese and Sierra)? Are body representation and internal simulation processes involved in coding a basic sense of self in artificial agents, and if

so, how (Schillaci et al.; Schillaci et al., 2016)? What should be built into an artificial agent "so that it really feels the touch of a finger, the redness of red, or the hurt of a pain" (O'Regan, 2014)? Terekhov and O'Regan show mathematically and in simulation that naive artificial agents can build the abstract notion of space from their perceptual systems by learning sensorimotor invariants. Without making assumptions about the existence of space, such agents are able to learn the notion of rigid displacement. Their findings give a role to artificial intelligence in the quest of explaining the nature of space, prevalently addressed by philosophy and physics.

Vernon et al. analyze the role of memory in anticipation processes. They propose a framework integrating procedural and episodic memory into internal simulation processes. Joint episodic and procedural memory facilitates prospection as it constrains the combinatorial explosion of potential perception-action associations allowing effective action selection in reaching goals.

2. EXPERIMENTAL STUDIES

Motor and cognitive developments in infancy are characterized by a process, where the individual is actively involved in the shaping of the experience through exploration (Schillaci et al.). Exposure to different sensorimotor experiences can influence cognitive development. Lones et al. demonstrate that this applies also to artificial agents. They observed that robots showed greater cognitive capabilities when exposed to a rich set of novel sensorimotor experiences as opposed to robots raised in poorly stimulant environments.

Exploration is the drive for motor and cognitive development (Schillaci et al.). The production of behavioral diversity is fundamental for discovering new aspects of the environment and of the individuals embodiment. Benureau and Oudeyer propose a mechanism for creating behavioral diversity in robots by re-enacting previously experienced sensorimotor actions. The artificial agent is capable of learning efficiently and adapting its past experience to new contexts even when characterized by high dimensional sensorimotor spaces. For example, by re-enacting past interaction with an object, the system is capable of learning more efficiently and capable of adapting to objects that are different in morphology.

How can this sensorimotor experience be stored in artificial agents? Vicente et al. propose a mechanism for simultaneous body schema adaptation and end-effector pose estimation on a humanoid robot. The system learns an internal body representation which is used to generate hypotheses of limb positions in space. These hypotheses are combined in a Bayesian fashion with the real perceptual feedback in a visual servoing control scheme that enables precise reaching actions.

Similarly, Escobar Juárez et al. address the development of internal body representations in artificial agents through Self-Organizing Maps and Hebbian learning. The authors present two experiments that show the capabilities of this architecture to implement internal simulation processes. Most importantly, these experiments show its potential as a building block for the coding of more complex sensorimotor schemes and behaviors. Schrodt and Butz propose a similar architecture, but with a probabilistic flavor. The system they propose is based on a stochastic generative neural network and is capable of implementing mental simulations in an artificial system. Moreover, the system learns to infer actions from partial sensory information and implements imagination capabilities to emulate and to recognize observed actions using the self body model.

Finally, Billing et al. present a learning mechanism that enables a simulated robot to mentally imagine navigating in an apartment environment. The proposed model can learn from human demonstrations and can re-enact the demonstrated behavior both overtly (using real sensory observations) and covertly (using simulated sensory observations).

3. CONCLUSION

The studies presented in this e-book offer support for the fundamental role that processes of re-enacting sensorimotor experience can have in the development of motor and cognitive skills, both in humans and in artificial agents. By bringing together theoretical review and experimental studies, we hope to further strengthen what we believe to be a very important research topic.

The works presented here study highly important but still very low-level processes. A fundamental issue to be addressed in future research is how the behavioral and computational components identified by the aforementioned studies support higher level cognition and action. In fact, the majority of the studies in the field address the bootstrapping of particular skills without explaining how the development of further skills may progress. We strongly believe that further investigation on the *interaction* between mechanisms for artificial curiosity, exploration, body representations, memory, and simulation processes would provide insights in the quest for open-ended development in artificial agents.

AUTHOR CONTRIBUTIONS

Each of the authors has contributed equally and significantly to the study.

FUNDING

REFERENCES

Case, L. K., Pineda, J., and Ramachandran, V. S. (2015). Common coding and dynamic interactions between observed, imagined, and experienced motor and somatosensory activity. *Neuropsychologia* 79, 233–245. doi:10.1016/j.neuropsychologia.2015.04.005

O'Regan, J. K. (2014). "The explanatory status of the sensorimotor approach to phenomenal consciousness, and its appeal to cognition," in *Contemporary Sensorimotor Theory*, eds J. M. Bishop, and A. O. Martin (Springer), 23–35.

Pezzulo, G., Candidi, M., Dindo, H., and Barca, L. (2013). Action simulation in the human brain: twelve questions. *New Ideas Psychol.* 31, 270–290. doi:10.1016/j.newideapsych.2013.01.004

Schillaci, G., Ritter, C. N., Hafner, V. V., and Lara, B. (2016). "Body representations for robot ego-noise modelling and prediction, towards the development of a sense of agency in artificial agents," in *International Conference on the Simulation and Synthesis of Living Systems (ALife XV)*, (Cancún), 390–397.

Wood, A., Rychlowska, M., Korb, S., and Niedenthal, P. (2016). Fashioning the face: sensorimotor simulation contributes to facial expression recognition. *Trends Cogn. Sci.* 20, 227–240. doi:10.1016/j.tics.2015.12.010

A Self-Organized Internal Models Architecture for Coding Sensory–Motor Schemes

Esau Escobar-Juárez[1], Guido Schillaci[2], Jorge Hermosillo-Valadez[1]* and Bruno Lara-Guzmán[1]

[1] Centro de Investigación en Ciencias-(IICBA), Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, México, [2] Adaptive Systems Group, Department of Computer Science, Humboldt-Universität zu Berlin, Berlin, Germany

Cognitive robotics research draws inspiration from theories and models on cognition, as conceived by neuroscience or cognitive psychology, to investigate biologically plausible computational models in artificial agents. In this field, the theoretical framework of Grounded Cognition provides epistemological and methodological grounds for the computational modeling of cognition. It has been stressed in the literature that *simulation*, *prediction*, and *multi-modal integration* are key aspects of cognition and that computational architectures capable of putting them into play in a biologically plausible way are a necessity. Research in this direction has brought extensive empirical evidence, suggesting that *Internal Models* are suitable mechanisms for sensory–motor integration. However, current Internal Models architectures show several drawbacks, mainly due to the lack of a unified substrate allowing for a true sensory–motor integration space, enabling flexible and scalable ways to model cognition under the embodiment hypothesis constraints. We propose the Self-Organized Internal Models Architecture (SOIMA), a computational cognitive architecture coded by means of a network of self-organized maps, implementing coupled internal models that allow modeling multi-modal sensory–motor schemes. Our approach addresses integrally the issues of current implementations of Internal Models. We discuss the design and features of the architecture, and provide empirical results on a humanoid robot that demonstrate the benefits and potentialities of the SOIMA concept for studying cognition in artificial agents.

Keywords: internal models, cognitive robotics, self-organized maps, sensory–motor schemes, computational architecture

1. INTRODUCTION

Cognitive robotics is an active research field in the cognitive sciences since the role of embodiment has been acknowledged as crucial to understand and reproduce natural cognition, showing as well a stance against the classic theory of cognition as symbolic processing. Research in cognitive robotics draws inspiration from theories and models on cognition, as conceived by neuroscience or cognitive psychology, to investigate biologically plausible computational models in artificial agents. Scientific aims include studying the implications of these models under controlled conditions and providing agents with basic cognitive skills (Pfeifer and Scheier, 2001).

In this research field, Grounded Cognition (Barsalou, 2008) constitutes a theoretical reference framework, including the account of embodied cognition, which stresses the importance of the body–environment interaction for the structuring and emergence of cognitive skills (Wilson, 2002).

Under this perspective, we are committed to investigate biologically plausible computational architectures in which to model cognition effectively. This issue has not trivial answers since there are several constraints on the nature, the role, and the architectural integration of the underlying artificial mechanisms by means of which we shall achieve computationally effective ways to model cognition. Some of these most relevant constraints are revised now.

In Grounded Cognition, all aspects of experience, perceptual states (for instance, those produced by vision, hearing, touching, tasting), together with internal bodily states and action, have neural correlates in the brain that are stored in memory. These neural activation patterns constitute multi-modal representations that are re-enacted during perception, memory, and reasoning. Modal re-enactments of these patterns constitute *internal simulation* processes (Barsalou, 2003) and are considered to lie at the heart of the off-line characteristics of cognition (Wilson, 2002). Thus, the theory of simulation is at the core of the embodied cognition hypothesis.

This shift in the paradigm about cognition has necessarily brought new design considerations on computer models in order to achieve embodied or grounded cognition, which have taken center stage in Artificial Intelligence [e.g., Grush (2004), Svensson and Ziemke (2004), and Pezzulo et al. (2011, 2013a)]. Thus, emphasis has been made on the predictive learning and internal modeling capabilities of the sensory–motor system (Pezzulo et al., 2013b).

Furthermore, in the embodied cognition framework, the acquisition of sensory–motor schemes is central (Pfeifer and Bongard, 2007), for they underlie cognition (Lungarella et al., 2003) and are grounded in the regularities of the sensory–motor system interactions with its environment. The cerebral cortex provides the necessary substrate for the development of these sensory–motor schemes as it constitutes the locus of the integration of multi-modal information.

All these considerations point to the fact that *simulation*, *prediction*, and *multi-modal integration* are key aspects of cognition and it has been stressed in the literature the necessity to achieve cognitive architectures capable of putting them into play in a biologically plausible way (Pezzulo et al., 2011). This paper is an attempt in this direction.

Based on extensive empirical evidence of its putative functionality in the Central Nervous System (Kawato, 1999; Blakemore et al., 2000; Wolpert et al., 2001), *Forward* and *Inverse Models* provide arguably a sound epistemological basis to understand cognitive processes at a certain level of description under the embodied cognition framework.

A thorough review of the implementations of internal models is out of the scope of this work. However, it is worth noting that most of the implementations show shortcomings in light of our previous discussion [e.g., see Arceo et al. (2013)]. First, we find that current implementations of internal models lack of flexibility as a consequence of the computational tools used. This translates into the fact that learning plasticity is highly reduced or even absent [e.g., see Lara and Rendon-Mancha (2006), Dearden (2008), Möller and Schenck (2008), and Schenck et al. (2011)]. Second, the implementations redound in *ad hoc* inverse and forward models, not easily scalable, and in some cases, using different networks for different motor commands [e.g., Möller and Schenck (2008)]. Finally, in the literature, there is an abstract and high-level coding of inverse models as in Dearden (2008), where inverse models are coded as direct actions.

We propose a new computational architecture for building cognitive tasks under the paradigm of Grounded Cognition: the Self-Organized Internal Models Architecture (SOIMA), a computational cognitive architecture coded by means of a network of self-organized maps, implementing coupled internal models that allow multi-modal associations. The SOIMA tackles integrally the issues of current implementations as will be discussed in the sequel.

The structure of the paper is as follows: in Section 2, we introduce the SOIMA architecture, explaining its theoretical foundations, justifying the Internal Models approach for modeling cognition and detailing the SOIMA's structure. Also, the features that make of it a suitable cognitive architecture tackling current implementations' shortcomings are introduced in this section. In Section 3, we provide two experimental case studies to demonstrate the architecture's functionality. We first introduce a case study for saccadic control in order to demonstrate the SOIMA features in detail. Then, we show a Hand–Eye Coordination task allowing us to demonstrate a scaling-up of the architecture, showing how the connectivity enhancements enable flexible and effective ways to model more complex tasks. In Section 4, we conclude by discussing the results and perspectives for future research.

2. SELF-ORGANIZED INTERNAL MODELS ARCHITECTURE: SOIMA

2.1. Biological Foundations

Brain plasticity regulates our capability to learn and to modify our behavior. Plastic changes are induced in neural pathways and synapses by the bodily experience with the external environment.

In the neurosciences literature, it has been proposed that the rich multi-modal information flowing through the sensory and motor streams is integrated in a sort of body schema, or body representation. Fundamental for action planning and for efficiently interacting with the environment (Hoffmann et al., 2010), such a body representation would be acquired and refined over time, already during pre-natal developmental stages.

For example, Rochat (1998) showed that infants exhibit, at the age of 3 months, systematic visual and proprioceptive self-exploration. The authors also report that infants, by the age of 12 months, possess a sense of a calibrated intermodal space of their body, that is a perceptually organized entity which they can monitor and control (Rochat and Morgan, 1998). As discussed by Maravita et al. (2003) and Maravita and Iriki (2004), converging evidence from animal and human studies suggests that the primate brain constructs various body-part-centered

representations of space, based on the integration of different motor and sensory signals, such as visual, tactile, and proprioceptive information.

Sensory receptors and effector systems seem to be organized into topographic maps that are precisely aligned both within and across modalities (Udin and Fawcett, 1988; Cang and Feldheim, 2013). Such topographic maps self-organize throughout the brain development in a way that adjacent regions process spatially close sensory parts of the body. Kaas (1997) reports a number of studies showing the existence of such maps in the visual, auditory, olfactory, and somatosensory systems, as well as in parts of the motor brain areas.

All this evidence suggests thus that cognition relies on self-organized body-mapping structures integrating sensory–motor information. But how does this integration takes place?

The work of Damasio (1989) and Meyer and Damasio (2009) proposes a functional framework for multi-modal integration supported by the theory of convergence-divergence zones (CDZ) of the cerebral cortex. This theory holds that specific cortical areas can act as sets of pointers to other areas and, therefore, relate various cortical networks to each other.

CDZ integrate low level cortical area networks (close to the sensory or motor modalities) with high level amodal constructs, which solves the problem of multi-modal integration since it enables the extraction of complex pure and non-segmented sensory information units and sensory–motor contingencies.

In the CDZ convergence process, modal information spreads to the multi-modal integration areas; while in the divergence process, multi-modal information propagates to modal networks generating the re-enactment of sensory or motor states. It is in this sense that the propagation bi-directionality provides the mechanism of mental imagery and the re-enactment of sensory–motor states.

This bi-directional capability is, thus, fundamental for multi-modal integration and, hence, for cognition. The lack of this property is precisely one of the main limitations of current cognitive architectures that our model tackles as will be shown in subsequent sections.

2.2. Internal Models

Internal models merge in a natural way sensory and motor information and create a multi-modal representation (Wolpert and Kawato, 1998). These models also provide agents with anticipation, prediction, and motor planning capabilities by means of internal simulations (Schillaci et al., 2012b).

We are particularly interested in the pair formed by Inverse-Forward models. The inverse model (IM) is a controller (**Figure 1A**), which generates the motor command (M_t^*) needed to achieve a desired sensory state (S_{t+1}) given a current sensory state (S_t). The forward model (FM) is a predictor (**Figure 1B**) that predicts the sensory state entailed (S_{t+1}^*) by some action of the agent (M_t) given a current sensory state (S_t).

While the IM is mainly required for motor control, the FM has been proposed as a possible model for a number of important issues, among which are sensory cancelation (Blakemore et al., 2000), state estimation (Wolpert et al., 1995), and body map acquisition (Schillaci et al., 2012a).

FIGURE 1 | Internal models canonical representation. (A) Inverse model: given the current sensory state St and a desired sensory state for time t + 1, St+1 this model suggests the motor command M_t^* that brings about the change in the sensory state of the agent. (B) Forward model: given the current sensory state S_t and a motor command M_t, the forward model predicts the resulting *sensory state prediction* S_{t+1}^*, were the command executed. It is worth noting the difference between S_{t+1}^* and S_{t+1}. The former is the actual expected state for the next time-step, but the latter may require several time-steps in order to be attained.

The coupled pair IM–FM has been introduced by Jordan and Rumelhart (1992) from control theory. In neuroscience, one of the first proposals was the MOSAIC architecture by Wolpert and Kawato (1998) and has been used in action recognition (Demiris and Khadhouri, 2006; Arceo et al., 2013), own body distinction (Schillaci et al., 2013), and mental simulation (Möller and Schenck, 2008).

In Cognitive Robotics, internal models have been used for action execution and recognition (Dearden, 2008), safe navigation planning (Lara and Rendon-Mancha, 2006; Möller and Schenck, 2008), and saccades control (Schenck et al., 2011). On the other hand, several works have proposed IM–FM couplings to perform different tasks. For example, in Schillaci et al. (2012b), several IM–FM pairs are used to recognize an action when comparing the output of each pair with the real situation. In the case of Schenck (2008), each pair is used in order to produce the motor command enabling an agent to reach a desired position, where the FM acts as the desired position monitor.

Internal Models are thus a suitable mechanism for multi-modal representations. They constitute a sound basis for modeling cognition and they also provide a coherent epistemological framework for studying it under the embodied cognition framework.

In our work, we propose an architecture that preserves the structural ideas put forward by Damasio along with the self-organizing and multi-modal integration properties of the brain, in the framework of internal models. This allows for building a mechanism for the integration and generation of multi-modal sensory–motor schemes in the framework of Grounded Cognition.

The SOIMA relies on two main learning mechanisms. The first one consists in Self-Organizing Maps (SOMs) that create clusters of modal information coming from the environment. The second one codes the internal models by means of connections between the first maps using Hebbian learning. This Hebbian association process generates sensory–motor patterns that represent actual sensory–motor schemes.

This coding approach of internal models using SOMs and Hebbian learning allows for a modular implementation, and constitutes the main contribution of our work. The architecture allows for an integrated learning strategy and provides means for building coupled sensory–motor schemes in a flexible way. Most of

the previous approaches of internal models implementations, as reported in the literature, provide different computational substrates that connect to each other in order to synthesize the coupled model. The substrates may even be of different nature (e.g., different kinds of neural networks), which obligates to use distinct learning strategies for each model. In many cases also, the resulting models are *ad hoc* for the task. Our approach synthesizes the coupled model on the same substrate, conferring connectivity enhancements that allow for modular and flexible internal models implementations and sensory–motor scheme maps. This mapping capability may be exploited in interesting ways as will be discussed in the conclusions.

We now discuss the implementation details of these two learning mechanisms in the SOIMA.

2.3. SOIMA's Structure

2.3.1. Modal Information Clustering

In the SOIMA, the basic units are SOMs (Kohonen, 1990) that generate clusters of information coming from different modalities (sensory or motor) or from other SOMs.

When training a SOM, a topological organization occurs in a space of lower dimension (2D or 3D) than the modal input space. This organization corresponds to a partition of the input space into regions that reveal the similarities of the input data.

A SOM is an artificial neural network endowed with an unsupervised learning mechanism based on vector quantization. Vector quantization refers to the fitness of a probability density function to a discrete set of prototype vectors.

In our case, we are interested in evaluating the differences between the vectors not only in terms of their relative distance but also in terms of their orientation. The cosine similarity can be used to obtain the differences in orientation, so we use both measures for clustering data in the SOM. Thus in our design, when a vector x occurs at the input, the activation A_j of each node in the SOM is defined as

$$A_j = \frac{1}{2}\left(\|\mathbf{x} - \mathbf{w}_j\| + 1 - \frac{\mathbf{x} \cdot \mathbf{w}_j}{\|\mathbf{x}\|\|\mathbf{w}_j\|} \right) \quad (1)$$

where $\mathbf{w}_j$ is the vector of weights between the input and the node j, the first term is the Euclidean distance between x and $\mathbf{w}_j$ and the second term is the cosine similarity. The winning node is the one with the lowest activation.

Once the winning node is computed, the weights of the neurons are updated according to the following equation

$$\Delta \mathbf{w}_j = \alpha(t) h_j (\mathbf{x} - \mathbf{w}_j) \quad (2)$$

where $\mathbf{w}_j$ is the weight between the input vector x and node j, h_j is the neighborhood function of node j defined as

$$h_j = e^{\left(\frac{-\beta_j}{2\sqrt{n}}\right)} \quad (3)$$

where β_j is the distance between node j and the winning unit, and n is the total number of nodes in the map. If β_j is greater than the size of the neighborhood v then $h_j = 0$, v decreases monotonically

from $v_i{}^1$ to $v_f = 1$, where v_i and v_f are the initial and final neighborhood sizes, respectively.

And finally $\alpha(t)$ is the learning rate that increases as a function of time, defined by:

$$\alpha(t) = \alpha^i + \left(\frac{t(\alpha^f - \alpha^i)}{(v_i - v_f)} \right) \quad (4)$$

where α^i and α^f are the initial and final learning rates, respectively, and t is the current learning period.

2.3.2. Modal Maps Association

The association between SOMs of different modalities has been reported in recent work. In Westerman and Miranda (2002), the association between vocalization and hearing maps can be used for modeling the emergence of vocal categories. In Li et al. (2007), clues about the vocabulary development age in infants, using a similar association scheme, were found. In Mayor and Plunkett (2010), an association between visual and hearing maps was used to determine the taxonomic response in early learning of words. Morse et al. (2010a) integrates different sensory and motor modal maps through a changing network with Hebbian learning to build a semantic meaning acquisition system.

In our work, we create the modal association between different SOMs through weights connected using the well-known Hebbian learning rule (Hebb, 1949). The rule states

> When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased (Hebb, 1949).

In this respect, the Hebbian rule establishes that the connection between neurons is reinforced according to the activation of neurons that participate in the connection. In our model, we use the following positive Hebbian rule for modulating connections between nodes of different maps:

$$\Delta w_{ij} = \alpha u_i u_j \quad (5)$$

where w_{ij} is the weight of the connection between the node the node i, and j α is the learning rate, u_i is the activation of the node i as u_j is the activation of the node j.

Figure 2 depicts the proposed architecture showing two maps (S and M) corresponding to the modalities of the agent. We consider M as a modality so that, together with S, the top map forms a multi-modal representation (*MMR*). The idea of Hebbian training is to modulate a network of connections between the top SOM and each modality SOM. Each node of the top SOM is connected to every node of the modality SOMs. These connections are originally set to 0.

The association process takes place as follows. The two maps S and M are fed with the sensory and motor data generated

---

[1]This initial value is oftentimes taken as the 15% of the size of the map.

FIGURE 2 | Self-organized internal models architecture (SOIMA).
Two modality maps are shown representing *M* and *S*. The map at the top
integrates the modalities coding, thus, sensory–motor schemes in a
Multi-Modal Representation (*MMR*).
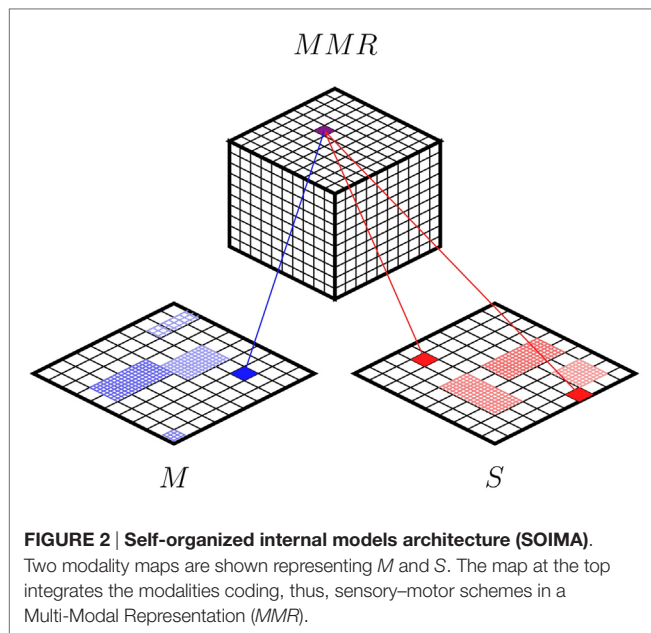
throughout the interaction of the agent with the environment.
Every time an input pattern is introduced, there is a winning
node in *S* and *M*, respectively. Then the inputs to the top map are
the coordinates of these winning units, so that a corresponding
winning unit occurs at the top SOM. A Hebbian modulation
is then applied to the connection between these nodes. In this
way, a sensory–motor scheme is coded on the *MMR* through the
Hebbian mechanism. Once the system has being trained, this
association allows for retrieving all the modalities when any of
them is present.

As can be seen in **Figure 2**, each winning unit of the *MMR*
map receives one connection coming from the *M* map and two
connections coming from the *S* map, representing two differ-
ent time steps (a change in the sensory situation); the motor
command is the one associated with that change in the sensory
situation. Thus, the trained system associates a triplet formed
by a sensory situation, a motor command, and a correspond-
ing predicted sensory situation associated with these two. It
could be said then that the *MMR* codes the associated triplet,
as each node in this map codes for a specific sensory–motor
experience of the agent. The *MMR* map is coded as a cube
in order to better represent the multi-modal space. In this
configuration, each triplet has 26 direct neighbors providing
a richer structure.

We now discuss the main attributes of the SOIMA.

2.4. SOIMA's Features
One of the main advantages of the SOIMA resides in its
bi-directional functionality, since it can work either as a forward
or as an inverse model, depending on the inputs that are fed to the
system. In other words, the system integrates the sensory–motor
scheme in such a way that it is now independent of the direction-
ality of the modal information flow.

When a forward model is required, an *S* and *M* signal for the
time *t* should be present, activating the corresponding maps and
their connections toward the *MMR* of the system[2] (see **Figure 2**).
Thence, the signal would spread back to the map *S*, producing with
its activation the prediction of the sensory state at the time $t + 1$.

If, on the contrary, an inverse model is needed, then the required
inputs are two sensory situations, coded in *S* corresponding to
times *t* and $t + 1$, in turn triggering the activation of the *MMR*
map and, thence, activating the node in *M* corresponding to the
pair in *S*.

Some interesting features of the *MMR* are noteworthy. The
MMR allows the bi-directional feature to be functional all
the time. In other words, any model (IM or FM) can be easily
implemented by instantiating the corresponding inputs for the
required functionality. In this way, both internal models are
coded on the same substrate, enabling the design of integrated
learning strategies.

Moreover, the *MMR* allows for the construction of coupled
IM–FM pairs in a modular way. Indeed, as either model (IM or
FM) can be instantiated at any time, the output of one of them
can be used as input to the other, constituting thus an IM–FM
coupling. Thus, several IM–FM couplings can be instantiated
sequentially, so as to build a simulation process for instance.
Hence, it is possible to feed sequentially in time the IM or FM,
either with data coming from the environment or produced
by the system itself. Also, the *MMR* allows for the integration
of other *MMR*s, which enables the coupling of distinct sen-
sory–motor schemes. These features will be illustrated in the
experiments.

Last, but not least, the *MMR* is not built from an abstract
representation, in the sense of being defined by the programmer
of the system as in the classical AI paradigm. Rather, it constitutes
a representation *grounded* in the bodily constrained interaction of
the agent with its surrounding environment.

The online learning ability of the architecture is also notewor-
thy. While running, the system is able to acquire new knowledge
in order to improve its performance on an ongoing basis. This
capability is achieved through the updating of the weights
interconnecting the SOMs, so the system can adapt to unfamiliar
situations as they arrive. This feature will be made clearer in the
next section.

3. EXPERIMENTS AND RESULTS

In order to demonstrate the functionality of the SOIMA, we
introduce two case studies using a NAO[3] humanoid robot. The
first experiment is intended to demonstrate the SOIMA's func-
tionality by modeling saccadic movements of the eye in order
to center a stimulus. This experiment was designed to describe
the detailed workings of the SOIMA implementation and shows
also that the SOIMA approach allows to cope with typical
dimensionality issues of the visual input space. The second case

---

[2]The activation of the top map constitutes the integration and multi-modal repre-
sentation of the event.
[3]Developed by Aldebaran Robotics.

study introduces an example of how the SOIMA can be scaled upwardly to model more complex cognitive tasks. In this second experiment, we aim at implementing a Hand–Eye Coordination (*HEC*) strategy using the SOIMA. Here, the architecture is used as a building block, allowing for exploiting previously acquired knowledge.

3.1. Saccadic Control
3.1.1. Visuomotor Schemes Modeling
Rapid eye motions, the so-called *saccadic movements* (Leigh and Zee, 1999), are intended to project the image of the visual area of interest in the most sensitive part of the retina, called the fovea. Saccadic control is a canonical problem involving sensory predictive and fine-tuning motor control capabilities.

It is worth mentioning a brief comment on the work of Kaiser et al. (2013) that introduces a saccadic control system based on internal models and asserts that addressing the image prediction problem is rather highly complex due to the input high dimensionality. For this reason, in their proposal predictions are not made, rather inverse mappings are computed from the output image to the input image.

The implementation of internal models can be made with any of the available learning techniques in artificial intelligence. However, there are serious limitations to the use of images, because they increase the dimensionality of the problem as well as the difficulty on finding regularities in the inputs. As an end result, the coding and learning of visuomotor schemes becomes a major difficulty.

By contrast, the scheme that we propose addresses the prediction problem using images of high dimensionality, which enables the development of more versatile models. Our system allows for learning the relationship between the camera motions and the corresponding sensory changes. Once learned, the model works as a mechanism that retrieves the motor command required to take some stimulus in the image to any desired area in the same image. In particular, we want the model to focus on some salient stimulus. A similar implementation is presented in Karaoguz et al. (2009) where a SOM is used for gaze fixation.

We use as input an image grabbed from one of the cameras in the NAO humanoid robot. Our experiment on saccadic control is an instance of a modular system that implements different coupled internal models pairs (FM–IM). We use the simulations provided by the FM–IM pair to provide the motor commands necessary to place some stimulus present in the image at any specified position. In **Figure 3,** we show the schematic functional diagram of the system.

At this point, it is worth mentioning that an important feature of the architecture is the ability to learn the Hebbian connections online. Initially, there is no association between the maps in the system, i.e., all connections have a value of 0; therefore, no motor command may be suggested aiming at focusing the S_t stimulus by means of the inverse model. Hence, when a motor command cannot be retrieved from the system, a motor-babbling mechanism generates a random movement. This command is then executed, obtaining thus the S_{t+1} image. This information is integrated into the connections between SOMs using on-line Hebbian learning.
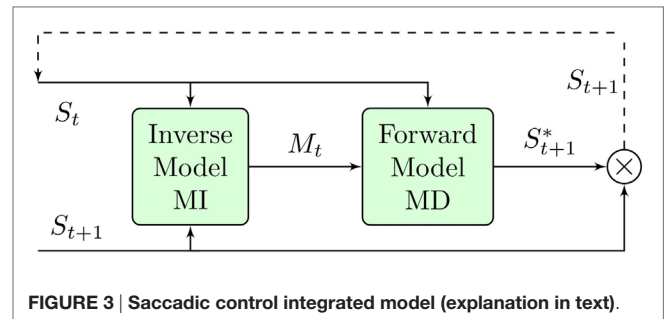
FIGURE 3 | Saccadic control integrated model (explanation in text).

The final result is the association of two nodes from the S map, one representing S_t and one representing S_{t+1} with one node in the M map representing the motor command that brings about this change in the sensory space.

Thus, in the first part of the forward-inverse coupling, input S_t represents the current visual sensory input, i.e., the image with a stimulus appearing at some arbitrary position. Input S_{t+1} represents the desired visual sensory state, i.e., the image with the stimulus in the desired position (this image is built or taken from a database). With these inputs, the inverse model suggests an initial motor command M_t^* aiming at the desired sensory change[4].

This motor command along with the image S_t is fed to the forward model to predict the sensory outcome S_{t+1}^*. This predicted image is compared with the desired one S_{t+1} to compute the error. In turn, the error is used to decide whether this output should be fed back as S_t, so that a corrective saccadic movement can be performed. In other words, supplementary control commands may be stacked together with the first M_t^* in order to reach the desired situation S_{t+1}.

Finally, once the motor commands required to bring S_t to the desired S_{t+1} are found, they are executed in the system with actual movements.

The details of the clustering of the modal maps and their Hebbian connections are discussed now.

Experiments were carried out on a simulated NAO humanoid robot endowed with 21 degrees of freedom (DOF) and two cameras. The Webots v8.0.3 was used to test the saccadic movements system. The experimental setup consisted in the robot facing a wall, situated at a distance of 40 cm, where a single visual stimulus was displayed, as shown in **Figure 4**. The model showed in **Figure 3** was used to control the saccadic movements toward the visual stimulus.

The two DOF associated to the agent's head movement were used: *yaw* (rotation around the vertical axis) and *pitch* (rotation around the horizontal axis); the camera is situated in the upper part of the head and has an image resolution of 640×480 pixels.

3.1.2. Sensory Input Processing
Learning requires a set of training patterns, each containing an image for S_t, one for S_{t+1} and a motor command that brings the

---

[4]Recall that once the system is trained, it can be used as an inverse or a forward model.

FIGURE 4 | Experimental setup.

FIGURE 5 | Input image processing, (A) original image 640 × 480, (B) foveated image 320 × 240, and (C) saliency image 40 × 30.
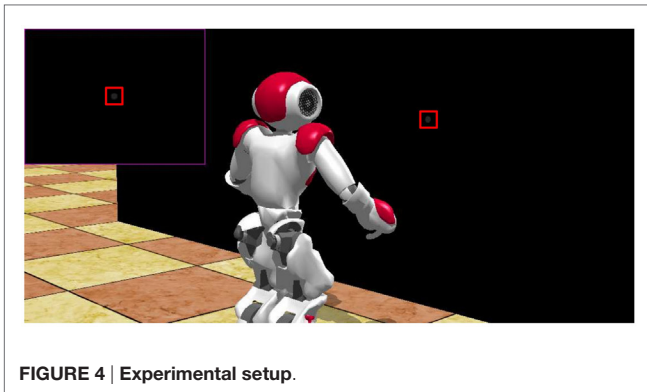
system from S_t to S_{t+1}. The motor command is defined as a pair (Δ Yaw, Δ Pitch).

To build the images, two stages of processing are necessary. In the first stage, a fovealized image is obtained from the original camera image (640 × 480 pixels). To foveate the images, we apply a radial mapping with high resolution toward the center of the image and low toward the periphery. The mapping emulates the human retina properties containing high concentration of photoreceptors in the fovea. It fulfills a special function in our design, because when a stimulus is located near the central part of the image (fovea) a small change in Yaw or Pitch corresponds to a large position change of the stimulus in the image. This enables a more accurate detection of the position of the stimulus nearby the central region of the image. As a result, the task of centering a stimulus in the image is more accurate. The fovealization algorithm delivers a 320 × 240 image.

The second processing phase is intended to facilitate the identification of *salient stimuli* in the image captured by the camera. This is achieved through binary thresholding and gaussian smoothing. At this stage, the image size is 40 × 30 pixels, reducing the sensory input space dimensionality. In **Figure 5,** we can see the visual stimuli at the different stages of processing.

In our system, a motor command M_t is a change (Δ) in the orientation (in degrees) on the horizontal and vertical axes of the robot's head. Any change between two positions depends on the resolution of the motors. This resolution in our case was built using a mapping mechanism similar to that applied to the visual modality. This motor mapping consisted in a variable yaw–pitch movement resolution, being this higher in the center than in the periphery.

To visualize the motor space, a motor resolution image (I_{RM}) was made from a total of 5000 head joints configurations. Initially, I_{RM} is an image where all its pixels are set to 0. Then for each position, the center of mass of the visual stimulus in the image is computed. According to the location of the center of mass, the value of the corresponding pixel in I_{RM} is increased. This gave rise to an intensity image where each pixel value is proportional to the number of positions in each location.

The I_{RM} image is depicted in **Figure 6**. As it can be seen, the highest visual stimulus density on camera positions is located in the central region of the image.
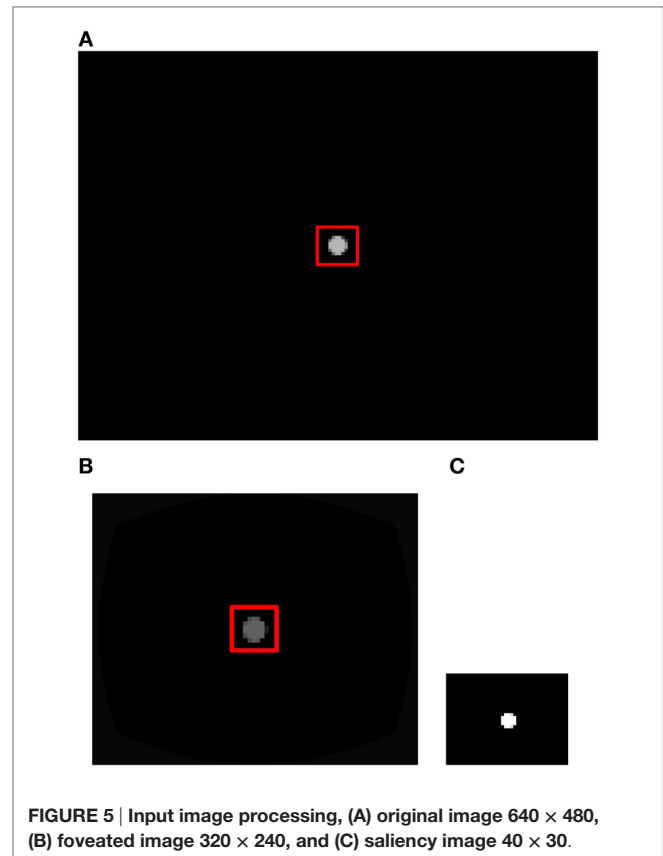
3.1.3 SOMs Training

For the training of the SOMs, 5000 random patterns with different initial (S_t) and final (S_{t+1}) camera positions and their corresponding M_t were taken with the following structure:

- S_t: it was formed by a 1200 values vector normalized from 0 to 1, taken from a 40 × 30 pixel salient features image.
- S_{t+1}: it was formed with the same procedure as for S_t but with a different camera position.
- M_t: it consists of two values (ΔYaw, ΔPitch) built in accordance with the following:

$$\Delta Yaw = (Yaw_{S_t} - Yaw_{S_{t+1}}) \tag{6}$$

$$\Delta Pitch = (Pitch_{S_t} - Pitch_{S_{t+1}}). \tag{7}$$

normalized from 0 to 1. With the robot placed at 40 cm from the stimulus, the Yaw axis of the camera covers an angle of 43.5° and the Pitch axis 37.8°, assuring that the stimulus is always on sight. The value of 1 represents the biggest positive possible change above the corresponding axis and 0 represents the biggest possible change in the other direction. Values of (0.5, 0.5) represent absence of movement in both axes.

The SOIMA used is shown in **Figure 2**. We used a 30 × 30 SOM to code S, and a 40 × 40 for M, finally for the *MMR* a three-dimensional 30 × 30 × 30 SOM was used. These maps were trained individually using the respective collected modal information patterns using the procedure described in Section 2.3.
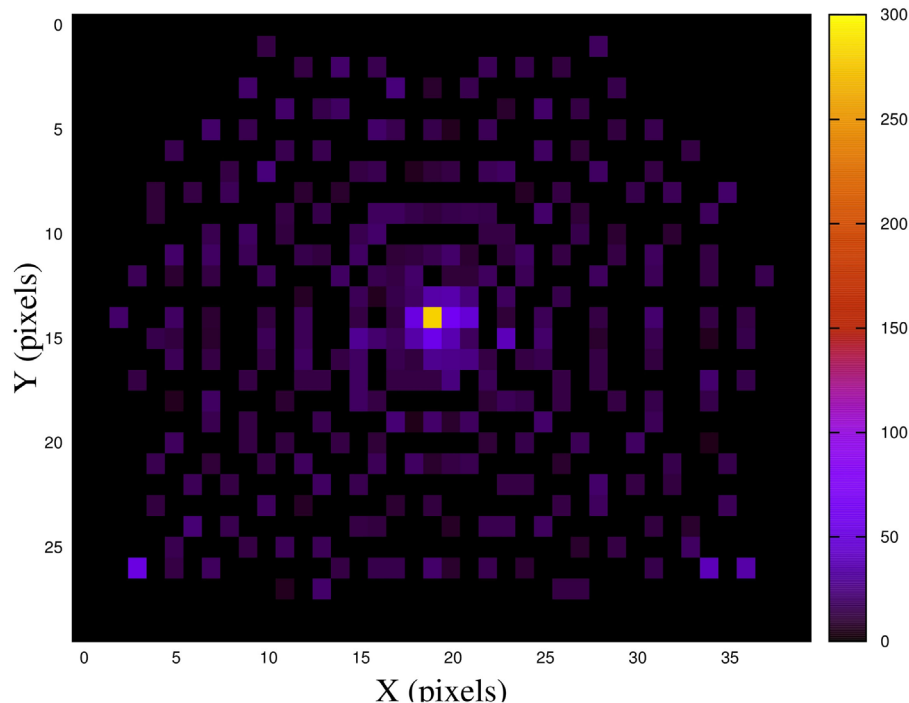
FIGURE 6 | **Visual stimulus distribution on different positions of the camera**.

3.1.4. Online Hebbian Learning

As mentioned in Section 3.1.1, the online learning capability gradually increases sensory–motor knowledge on saccadic movements, which decreases the use of the random mechanism (**Figure 7A**). In turn, the system reduces the error as it focuses the visual stimulus (**Figure 7B**).

Figure 7B depicts the quartiles of 11 subsets of the available data on the online learning error. Red lines show the medians of each subset. The figure shows that the variability of the error reduces with time. The mean value of each subset stabilizes quickly and falls within the third quartile, showing that the distributions of these data sets are not gaussian.

When the motor command is suggested by the system (i.e., when the SOIMA already contains a multi-modal association and is able to act as IM and FM), learning also occurs:

- when the architecture generates a sequence of two or more simulated saccadic movements to reach from S_t to S_{t+1}, a single motor command is learned as the association between the two sensory situations.
- in those cases, when after running a motor command the position error is greater than a certain threshold, another motor command is calculated and executed by re-enacting the system with S_{t+1}^* [5]. This new motor command is added to

the previous motor command, so that this new association is integrated into the system as if it were a single execution.

3.1.5. Saccadic Control Execution
3.1.5.1. Prediction
For illustration purposes, we want the system to center the stimulus on the image in a foveation-like process.

In principle, training data covers the whole of the visual field of the camera, so it would be possible to give another location of the stimulus as desired sensory situation S_{t+1}. However, the desired stimulus S_{t+1} is built from a repository of images containing a single object in the center of the image.

A typical test example is depicted in **Figure 8**, in this two prediction steps are conducted:

- Step 1: the current camera image is fed into the system and processed according to areas of salient features to form S_t. The desired image S_{t+1} is also fed into, with the stimulus located at the center. These inputs go through the inverse model and generate a suggestion for the motor command M_t^*. The latter is in turn applied together with S_t to the forward model to generate a prediction S_{t+1}^*. This output is compared with the S_{t+1}, the desired situation, to compute the error.
- Step 2: if the error is greater than a threshold, S_{t+1}^* is fed back into the system again, as the new S_t, to obtain an additional motor command in order to achieve the marked position accurately.

It is worth noting that for both steps, the error is calculated between the prediction of the forward model and the desired

---

[5]The reader must bear in mind that the system codes the FM–IM coupling. In other words, it is possible to generate new motor commands from the FM predicted sensory output (S_{t+1}^*), by using it as input again to the IM. This new motor command can in turn be used once again to generate a new prediction.

FIGURE 7 | Performance of the system during training. (A) Average percentage of random motor commands used as learning occurs. **(B)** Error in pixels between final position and center of image.
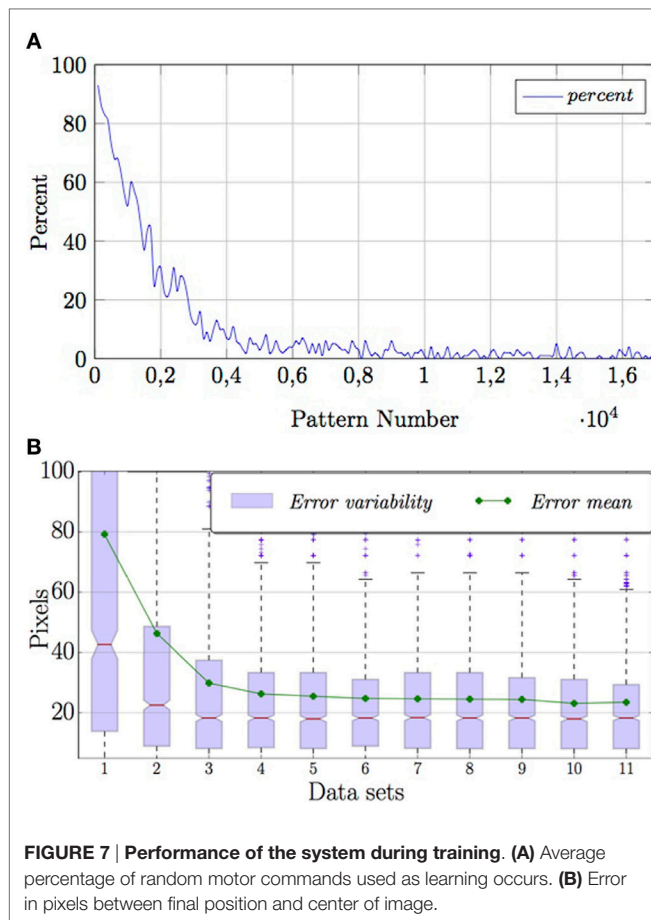
situation. Only when the error between the prediction and the desired state is lower than the threshold, a motor command is executed. This could mean that more than two internal simulations are run.

Ninety-five tests were performed on the saccadic control model (**Figure 9**) from different robot's head positions, with the initial stimulus located at distinct locations on the captured camera image (blue squares). After two internal simulations, the position of the stimulus is shown with red crosses with a mean error of 35.72 pixels (red disk), which means a 1.3% of the total image size and 3° of robot's motion.

The error coming from the saccadic movements are chiefly due to the visual fovealization since it reduces the information available around the image periphery, which causes precision loss in determining the initial stimulus location.

3.1.5.2. Execution
It is known that two components of a motoneuronal control signal generate the saccadic eye movement in humans (Bahill et al., 1975). These components correspond to an initial saccade and a corrective saccade.

Based on this fact, after realizing the sequence of moves suggested by the model (initial or approaching saccade), a second tuning or corrective motion is executed (see **Figure 10**). These two execution steps are described next.

- *Approaching saccadic movement:* the current image of the camera, S_t, and the target image S_{t+1}, with the stimulus at the center, are fed into the system. These inputs go through SOIMA and generate a suggested motor command M_t^* that is applied to

FIGURE 8 | Typical prediction example.
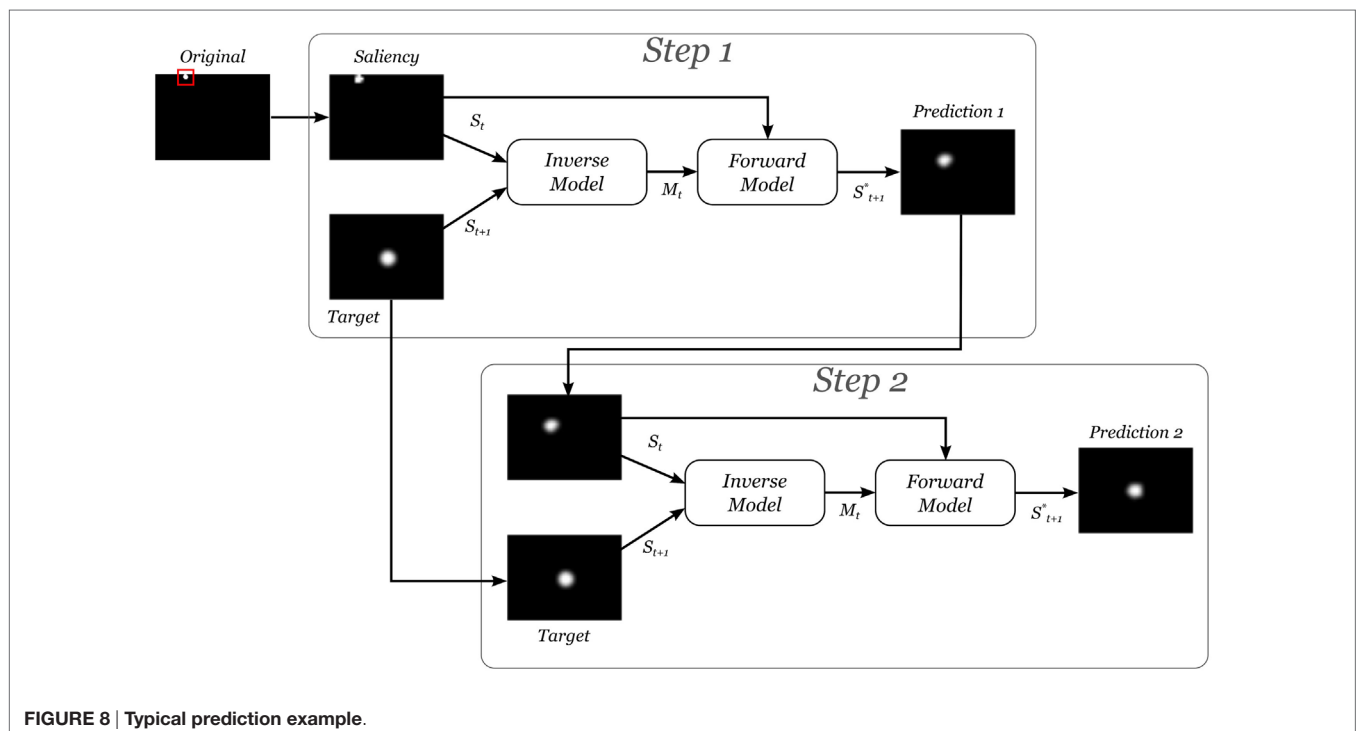
the Nao Robot. The error of the resulting stimulus location in the picture is calculated in order to know whether a second movement, for better accuracy, is necessary.

- *Tuning saccadic movement:* if the error is greater than a threshold of 10 pixels (0.1% of image surface and 0.8° of robot's movement), the actual image is fed back again to SOIMA to obtain an additional tuning motor command in order to reach the desired position accurately.

As opposed to the control described in Section 3.1.5, in this case motor commands are actually executed in both movements.
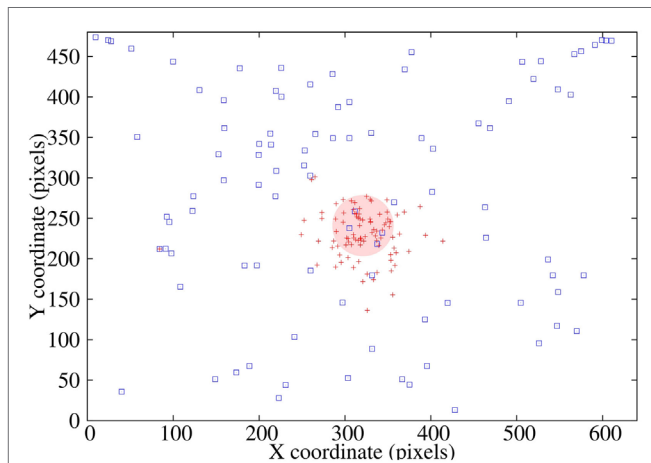
FIGURE 9 | Prediction performance: position of the stimulus after two predictions for 95 random initial positions.

In practice, this means that both movements in this strategy could contain more than one simulation step.

The system was tested on 84 patterns and it was found an average error of 36.1 pixels on the original 640 × 480 pixels image, which corresponds to a 3.1° error on approaching saccadic movements. For the tuning motion, a mean error of 19.3 pixels was obtained corresponding to a 1.6° error (see **Figure 11**).

3.1.5.3. Tracking of Stimulus
In addition, we realized a tracking experiment. The stimulus was moved around the wall plane facing the robot, while the latter executed a centering or foveation task by means of the acquired saccadic control model. The purpose of this experiment is to show that even when the stimulus spatial reference with respect to the robot changes, and so do the perspective, the agent is able to effectively use the saccadic controller.

In **Figure 12,** we depict a path following task. Red arrows show the sequence of the path followed by the stimulus. Numbers associated with each arrow correspond to the saccadic movements executed to center the stimulus on the image. Finally, blue numbers and circles show the centering task error in pixels over every point of the path.

3.2. Hand–Eye Coordination
Coordination of visual perception with body movements is an important prerequisite for the development of complex motor abilities.

Visuomotor coordination refers to the process of mapping visuospatial information into patterns of muscular activation. Such mapping is learned through the interaction of the agent with
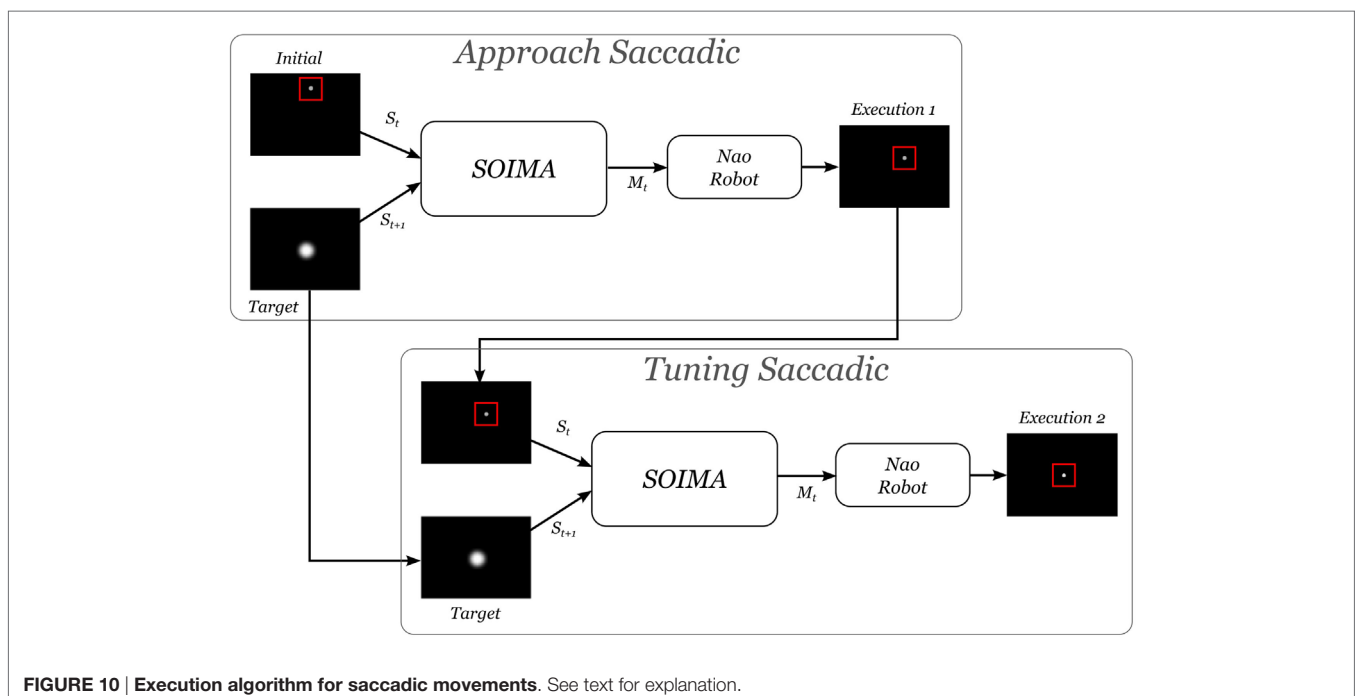
FIGURE 10 | Execution algorithm for saccadic movements. See text for explanation.

FIGURE 11 | Execution performance, including coarse and fine motions: initial stimulus locations in the picture through 84 test patterns (blue squares), ending stimulus locations during the approaching saccadic movement (red crosses), mean error (red circle), and final stimulus locations during the tuning saccadic movement (green crosses).

FIGURE 12 | Path followed by the stimulus on the robot's image plane. Red arrows and numbers show the path sequence and the number of saccadic movements in each centering task. Blue circles and numbers depict the error in pixels after centering the stimulus.

his environment. In particular Hand–Eye Coordination (*HEC*) refers to the coordinated use of the eyes with one or both hands to perform a task.

Here, we propose an implementation of the SOIMA for the learning of a sensory–motor scheme that allows *HEC* in a NAO humanoid robot. Once this coordination is learned then, given a particular posture (i.e., arm and hand postures) of the robot, the system should provide a head posture such that the hand appears in the visual field. It is worth mentioning that a posture is determined by absolute joint angles.

The SOIMA structure proposed can be seen in **Figure 13**, showing the integration of a *HEC* Multi-Modal Representation,

together with the saccadic control presented in the previous section.

Here, *V* is an 80 × 80 SOM coding the coordinates of the position of the robot's hand in the image plane. The image was obtained from the lower camera in the head of the robot. The coordinates were estimated with the use of the ARToolkit library,[6] using a fiduciary marker in the hand of the robot.

Head is a 100 × 100 SOM coding the values of the two degrees of freedom of the head *(yaw and pitch)*. *Arm* is a 80 × 80 SOM coding *shoulder pitch, shoulder roll, elbow yaw, and elbow roll* of the left robot arm. Finally, MMR_{eh} is a 150 × 150 SOM that codes the Multi-modal Representation of the sensory–motor scheme. The SOMs were trained using 6453 random patterns. The saccadic control used the same SOMs described in the previous experiment.

In this experiment, the system for saccadic control was used only as a tool for the training and testing of the *HEC* system. Given that both the *HEC* and the saccadic control system use the same visual input, the map *V* was the same as previously defined.

The experiments were carried out using the simulated NAO in an empty arena as shown in **Figure 14**.

Motor babbling was used in order to train the Hebbian associations for the *HEC* system. The general procedure for training was:

- Execution of random head and arm movements.
- Verify whether the marker was detected. When detected, the positions of head and arm as well as the visual information were fed into the system.
- The connection between the winning units in each of the three maps was reinforced.

During training, the saccadic control system was used to increase the variability and precision of the patterns used for the Hebbian associations. In the cases where the marker was found in the image, but not in the fovea, the saccadic control system was used to center the stimulus.

Two tests were carried out to assess the full system:

- First, the robot performed random arm movements and these were fed into the system; the head would then follow successfully the hand position, centering the stimulus in the foveal area.
- Second, a pointing test was implemented. A marker was randomly placed in the arena, the system would then perform a random exploration of the visual space. Once the maker was in sight, the saccadic control system would center the stimulus. The *HEC* system would then successfully position the hand before the marker.

Video material on both tests is available at the following url links: Test 1 on simulated robot[7]; Test 1 on real robot[8]; Test 2 on simulated robot[9]; Test 2 on real robot.[10]

---

[6]www.hitl.washington.edu/artoolkit

[7]https://youtu.be/agrkeUxiQZA

[8]https://youtu.be/7h_luKEre5s

[9]https://www.youtube.com/watch?v=BrFS7EWz4kc

[10]https://www.youtube.com/watch?v=ed7WkMgjybo
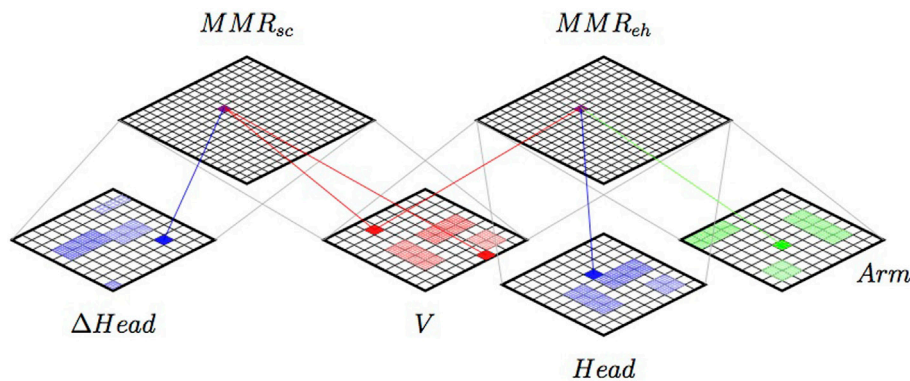
FIGURE 13 | The SOIMA up-scaled when implementing a Hand–Eye Coordination strategy, including the saccadic control presented in the previous section. Two Multi-Modal Representations (sensory–motor schemes) are coupled together through modal maps. See explanation in text.
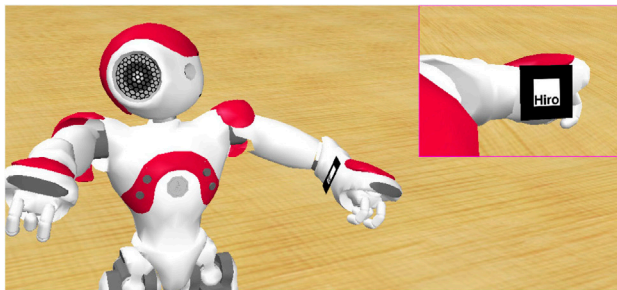
FIGURE 14 | Hand–Eye Coordination task experimental setup. Given particular arm and hand postures of the robot, the system provides a head posture such that the hand appears in the visual field.

4. DISCUSSION AND CONCLUSION

The relevance of modeling sensory–motor schemes relies on the fact that they are considered to be the fundamental unit of analysis for cognitive processes and skills under the cognitive robotics school of thought (Lungarella et al., 2003). Cognition relies on self-organized structures integrating sensory–motor information. As internal models create naturally a multi-modal representation of sensory–motor flows, they have been extensively studied as a suitable mechanism for sensory–motor integration.

Based on these considerations, we developed a new computational architecture called SOIMA, drawing biological inspiration from the theory of convergence–divergence zones of the cerebral cortex proposed by Damasio (1989) and Meyer and Damasio (2009) and from the self-organizing properties of the brain.

In order to introduce the architecture and prove its feasibility and performance, we implemented two case studies. The first experiment implemented a strategy of saccadic movement control consisting in centering a salient stimulus in the visual sensory space using the SOIMA. This experiment shown that the SOIMA approach allows to cope with vision issues regarding the input space dimensionality. The second case study implemented a Hand–Eye Coordination strategy allowing to show how the

SOIMA can be scaled upwardly in order to model more complex cognitive tasks.

The SOIMA integrates important qualities of online learning and introduces a novel form of internal models implementation not reported before. Even though there exists work showing coupled Self-Organized Networks (Hikita et al., 2008; Luciw and Weng, 2010; Morse et al., 2010b; Lallee and Dominey, 2013), our proposal goes a step further in that we model the predictive capabilities of the human cognitive machinery by means of internal models. However, current internal models architectures show major drawbacks so as to model cognition under the embodiment hypothesis constraints (e.g., independent coding of the inverse and forward models). The main attributes of the SOIMA provide means for autonomous sensory–motor integration, as it allows for multi-modal activation patterns to organize themselves into a coherent structure through Hebbian association, creating thus a multi-modal grounded representation. The bi-directional capability of the SOIMA allows this representation to become a sensory–motor scheme available as both a forward (predictive) and an inverse (controlling) model. The lack of this property is precisely one of the main limitations of current cognitive architectures. This bi-directional mechanism provides, thus, a unified substrate allowing for a true sensory–motor integration space and for coherent sensory–motor learning strategies.

We would like to highlight five main features making the SOIMA stand apart from current implementations of internal models and sensory–motor schemes. The first three features have been tested in the case studies presented here, the remaining two represent current work:

- *Modularity and scalability*: the experiments reported here exemplify the modular character of SOIMA. This feature redounds in an integrated learning strategy and allows for the scalability of the system. The architecture is modular in that the logical structure of the *MMR* is not hardwired, but develops as the agent interacts with the environment. This in turn provides means for the construction of sensory–motor schemes that can be sequentially re-enacted to accomplish a particular task. The workings of the architecture enable the system to learn

online both, the FM and the IM, in an integrated way. New examples of sensory–motor schemes are acquired as the agent experiences the world; thus, incorporating new knowledge for later use. A first example of the scalability of the system was reported here. Every new sensory–motor scheme generates a new *MMR*, coding a particular IM–FM coupling. Thus, different sensory–motor schemes can be coupled together in order to increase the sensory–motor capabilities of the agent. In this sense, the system is scalable. In summary, the SOIMA should be seen as a core unit for building more complex structures allowing to re-use previous knowledge.

- *Bidirectionality*: given the internal coding of the relations between the sensory and motor modalities, the SOIMA works as either a Forward or an Inverse model. The connections between the maps admit the bidirectional flow of information, therefore choosing the model depends on the question asked and the available input. As mentioned before, current implementations use mostly MLP networks, with one or various networks coding the Forward model and separately networks coding the Inverse model, which obligates to use different learning strategies for each model. Our approach synthesizes the coupled model on the same substrate, conferring connectivity enhancements that allow for integrated learning strategies and sensory–motor scheme maps.

- *Temporality*: different moments in time for the sensory situation are coded in the same map. The temporal relations between situations are coded in the Hebbian connections between maps. As a consequence, several time steps can be integrated into the same sensory–motor scheme. Another concern regarding temporality is stability. That is, whether the system will be able to cope with environmental changing conditions, i.e., become stable after some perturbation. This is indeed a major concern that is currently being investigated. Future work includes experimenting with other kinds of self-organizing maps (e.g., Dynamic SOMs). It is expected that these other maps would also allow for sensory–motor scheme reconfiguration in the long run, if ever the available

knowledge is not enough to properly model contingent task changes.

- *Motor Mapping*: the characterization of the information in the motor map should allow for trajectory generation in the motor space. Moving in the motor map from unit to unit would then have a mapping in the physical space of the agent.

- *Robustness to lack of information*: the structure proposed can be the base for capabilities such as action recognition. The agent codes a SOIMA based on its own experience and its own sensory–motor model; however, when observing the execution of an action some information would not be available (i.e., proprioceptive information). The lack of this information should not represent a problem as the activation produced by the available input should propagate the activation in the rest of the architecture.

The results presented here are encouraging and permit us to assert that the organization and functioning of the SOIMA is promising for undertaking research in further directions. In the context of Grounded Cognition, we consider that our work constitutes a biologically plausible computational approach, effective for the development of complex cognitive behavior models. As such, we hope that the SOIMA concept will enable the study and test of diverse hypotheses on the underpinning processes of cognition and the development of artificial agents exhibiting coherent behavior in their environment.

AUTHOR CONTRIBUTIONS

EE, JH, and BL designed research. EE performed research. EE, GS, JH, and BL wrote the paper.

ACKNOWLEDGMENTS

REFERENCES

Arceo, D. C., Escobar, E., Hermosillo, J., and Lara, B. (2013). Model of a mirror neuron system in an artificial autonomous agent. *Nova Scientia* 5, 51–72.

Bahill, A. T., Clark, M. R., and Stark, L. (1975). Glissades – eye movements generated by mismatched components of the saccadic motoneuronal control signal. *Math. Biosci.* 26, 303–318. doi:10.1016/0025-5564(75)90018-8

Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 1177–1187. doi:10.1098/rstb.2003.1319

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi:10.1146/annurev.psych.59.103006.093639

Blakemore, S.-J., Wolpert, D., and Frith, C. (2000). Why can't you tickle yourself? *Neuroreport* 11, R11–R16. doi:10.1097/00001756-200008030-00002

Cang, J., and Feldheim, D. A. (2013). Developmental mechanisms of topographic map formation and alignment. *Annu. Rev. Neurosci.* 36, 51–77. doi:10.1146/annurev-neuro-062012-170341

Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Comput.* 1, 123–132. doi:10.1162/neco.1989.1.1.123

Dearden, A. (2008). *Developmental Learning of Internal Models for Robotics*. Ph.D. thesis, Imperial College London, London.

Demiris, Y., and Khadhouri, B. (2006). Hierarchical attentive multiple models for execution and recognition of actions. *Rob. Auton. Syst.* 54, 361–369. doi:10.1016/j.robot.2006.02.003

Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behav. Brain Sci.* 27, 377–396. doi:10.1017/S0140525X04000093

Hebb, D. (1949). *The Organization of Behavior A Neuropsychological Theory*. New York: Lawrence Erlbaum Associates.

Hikita, M., Fuke, S., Ogino, M., Minato, T., and Asada, M. (2008). "Visual attention by saliency leads cross-modal body representation," in *Development and Learning, 2008. ICDL 2008. 7th IEEE International Conference on* (Monterey, CA: IEEE), 157–162.

Hoffmann, M., Marques, H., Hernandez Arieta, A., Sumioka, H., Lungarella, M., and Pfeifer, R. (2010). Body schema in robotics: a review. *IEEE Trans. Auton. Ment. Dev.* 2, 304–324. doi:10.1109/TAMD.2010.2086454

Jordan, M. I., and Rumelhart, D. E. (1992). Forward models: supervised learning with a distal teacher. *Cogn. Sci.* 16, 307–354. doi:10.1207/s15516709cog1603_1

Kaas, J. H. (1997). Topographic maps are fundamental to sensory processing. *Brain Res. Bull.* 44, 107–112. doi:10.1016/S0361-9230(97)00094-4

Kaiser, A., Schenck, W., and Möller, R. (2013). Solving the correspondence problem in stereo vision by internal simulation. *Adapt. Behav.* 21, 239–250. doi:10.1177/1059712313488425

Karaoguz, C., Dunn, M., Rodemann, T., and Goerick, C. (2009). "Online adaptation of gaze fixation for a stereo-vergence system with foveated vision," in *Advanced Robotics, 2009. ICAR 2009. International Conference on* (Munich: IEEE), 1–6.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol.* 9, 718–727. doi:10.1016/S0959-4388(99)00028-8

Kohonen, T. (1990). The self-organizing map. *Proc. IEEE* 78, 1464–1480. doi:10.1109/5.58325

Lallee, S., and Dominey, P. F. (2013). Multi-modal convergence maps: from body schema and self-representation to mental imagery. *Adapt. Behav.* 21, 274–285. doi:10.1177/1059712313488423

Lara, B., and Rendon-Mancha, J. M. (2006). "Prediction of undesired situations based on multi-modal representations," in *Electronics, Robotics and Automotive Mechanics Conference, 2006*, Vol. 1 (Cuernavaca: IEEE), 131–136.

Leigh, R. J., and Zee, D. S. (1999). *The Neurology of Eye Movements*, 3rd Edn. Oxford: Oxford University Press.

Li, P., Zhao, X., and Mac Whinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cogn. Sci.* 31, 581–612. doi:10.1080/15326900701399905

Luciw, M., and Weng, J. (2010). Top-down connections in self-organizing hebbian networks: topographic class grouping. *IEEE Trans. Auton. Ment. Dev.* 2, 248–261. doi:10.1109/TAMD.2010.2072150

Lungarella, M., Metta, G., Pfeifer, R., and Sandini, G. (2003). Developmental robotics: a survey. *Connect. Sci.* 15, 151–190. doi:10.1080/09540090310001655110

Maravita, A., and Iriki, A. (2004). Tools for the body (schema). *Trends Cogn. Sci.* 8, 79–86. doi:10.1016/j.tics.2003.12.008

Maravita, A., Spence, C., and Driver, J. (2003). Multisensory integration and the body schema: close to hand and within reach. *Curr. Biol.* 13, R531–R539. doi:10.1016/S0960-9822(03)00449-4

Mayor, J., and Plunkett, K. (2010). A neurocomputational account of taxonomic responding and fast mapping in early word learning. *Psychol. Rev.* 117, 1. doi:10.1037/a0018130

Meyer, K., and Damasio, A. (2009). Convergence and divergence in a neural architecture for recognition and memory. *Trends Neurosci.* 32, 376–382. doi:10.1016/j.tins.2009.04.002

Möller, R., and Schenck, W. (2008). Bootstrapping cognition from behavior – a computerized thought experiment. *Cogn. Sci.* 32, 504–542. doi:10.1080/03640210802035241

Morse, A. F., Belpaeme, T., Cangelosi, A., and Smith, L. B. (2010a). "Thinking with your body: modelling spatial biases in categorization using a real humanoid robot," in *Proc. of 2010 Annual Meeting of the Cognitive Science Society* (Portland: Cognitive Science Society), 1362–1368. Available at: http://www.proceedings.com/09137.html

Morse, A. F., De Greeff, J., Belpeame, T., and Cangelosi, A. (2010b). Epigenetic robotics architecture (era). *IEEE Trans. Auton. Ment. Dev.* 2, 325–339. doi:10.1109/TAMD.2010.2087020

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., and Spivey, M. J. (2013a). Computational grounded cognition: a new alliance between grounded cognition and computational modeling. *Front. Psychol.* 3:612. doi:10.3389/fpsyg.2012.00612

Pezzulo, G., Candidi, M., Dindo, H., and Barca, L. (2013b). Action simulation in the human brain: twelve questions. *New Ideas Psychol.* 31, 270–290. doi:10.1016/j.newideapsych.2013.01.004

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., Spivey, M., and McRae, K. (2011). The mechanics of embodiment: a dialogue on embodiment and computational modeling. *Front. Psychol.* 2:5. doi:10.3389/fpsyg.2011.00005

Pfeifer, R., and Bongard, J. (2007). *How the Body Shapes the Way We Think: A New View of Intelligence*. Cambridge: MIT press.

Pfeifer, R., and Scheier, C. (2001). *Understanding Intelligence*. Cambridge: MIT press.

Rochat, P. (1998). Self-perception and action in infancy. *Exp. Brain Res.* 123, 102–109. doi:10.1007/s002210050550

Rochat, P., and Morgan, R. (1998). Two functional orientations of self-exploration in infancy. *Br. J. Dev. Psychol.* 16, 139–154. doi:10.1111/j.2044-835X.1998.tb00914.x

Schenck, W. (2008). *Adaptive Internal Models for Motor Control and Visual Prediction. Number 20*. Berlin: Logos Verlag Berlin GmbH.

Schenck, W., Hoffmann, H., and Möller, R. (2011). Grasping of extrafoveal targets: a robotic model. *New Ideas Psychol.* 29, 235–259. doi:10.1016/j.newideapsych.2009.07.005

Schillaci, G., Hafner, V. V., and Lara, B. (2012a). "Coupled inverse-forward models for action execution leading to tool-use in a humanoid robot," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction* (Boston, MA: ACM), 231–232.

Schillaci, G., Lara, B., and Hafner, V. (2012b). "Internal simulations for behaviour selection and recognition," in *Human Behavior Understanding, Volume 7559 of Lecture Notes in Computer Science* (Berlin: Springer), 148–160.

Schillaci, G., Hafner, V. V., Lara, B., and Grosjean, M. (2013). "Is that me?: sensorimotor learning and self-other distinction in robotics," in *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction* (Tokyo: IEEE Press), 223–224.

Svensson, H., and Ziemke, T. (2004). "Making sense of embodiment: simulation theories and the sharing of neural circuitry between sensorimotor and cognitive processes," in *Presented at the 26th Annual Cognitive Science Society Conference* (Chicago, IL).

Udin, S. B., and Fawcett, J. W. (1988). Formation of topographic maps. *Annu. Rev. Neurosci.* 11, 289–327. doi:10.1146/annurev.ne.11.030188.001445

Westerman, G., and Miranda, E. R. (2002). Modelling the development of mirror neurons for auditory-motor integration. *J. New Music Res.* 31, 367–375. doi:10.1076/jnmr.31.4.367.14166

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636. doi:10.3758/BF03196322

Wolpert, D. M., Ghahramani, Z., and Flanagan, J. R. (2001). Perspectives and problems in motor learning. *Trends Cogn. Sci.* 5, 487–494. doi:10.1016/S1364-6613(00)01773-3

Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science* 269, 1880–1882.

Wolpert, D. M., and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Netw.* 11, 1317–1329. doi:10.1016/S0893-6080(98)00066-5

From Sensorimotor Experiences to Cognitive Development: Investigating the Influence of Experiential Diversity on the Development of an Epigenetic Robot

John Lones*, Matthew Lewis and Lola Cañamero

Embodied Emotion, Cognition and (Inter-)Action Lab, School of Computer Science, University of Hertfordshire, Hatfield, UK

Using an epigenetic model, in this paper we investigate the importance of sensorimotor experiences and environmental conditions in the emergence of more advanced cognitive abilities in an autonomous robot. We let the robot develop in three environments affording very different (physical and social) sensorimotor experiences: a "normal," standard environment, with reasonable opportunities for stimulation, a "novel" environment that offers many novel experiences, and a "sensory deprived" environment where the robot has very limited chances to interact. We then (a) assess how these different experiences influence and change the robot's ongoing development and behavior; (b) compare the said development to the different sensorimotor stages that infants go through; and (c) finally, after each "baby" robot has had time to develop in its environment, we recreate and assess its cognitive abilities using different well-known tests used in developmental psychology such as the violation of expectation (VOE) paradigm. Although our model was not explicitly designed following Piaget's or any other developmental theory, we observed, and discuss in the paper, that relevant sensorimotor experiences, or the lack of, result in the robot going through unforeseen developmental "stages" bearing some similarities to infant development, and could be interpreted in terms of Piaget's theory.

Keywords: epigenetic development, developmental robotics, sensorimotor development, cognitive development, social robotics, affective adaptation, human–robot interaction, autonomous robot

1. INTRODUCTION

The first 2 years of life represent a period of rapid cognitive development in human infants. During this 2-year period, behavioral patterns shift from simple reactions to incorporating the use of symbols in mental representations, setting the stage for further cognitive development (Piaget, 1952; DeLoache, 2000). While this cognitive development process is still not fully understood, some evidence does suggest that early stimulation provides the foundation for this process (Piaget, 1952; Fischer, 1980; Fuster, 2002; Bahrick et al., 2004). This theory of development is perhaps best explained in Piaget's concept of sensorimotor development. According to Piaget's work (Piaget, 1952), during the sensorimotor period infants go through 6 substages of development, which we will briefly lay out.

The first stage, often simply referred to as the "reflex stage," lasts from birth until around 1 month of age, with infants limited to simple automatic "innate" behaviors (Piaget and Inhelder, 1969). The second stage, known as "primary circular reactions," occurs approximately between 1 and 4 months and sees the infant's behavior begin to incorporate, repeat, and refine reflex behaviors focused on their own bodies. In the third stage, which takes place between the 4th and 8th month, infants start to notice that their actions can have interesting effects on their immediate environment. By around the 8th to 12th month, infants begin to display coordination in their secondary circular reactions facilitating goal-directed behavior. In addition, during this stage, infants begin to show signs of understanding the concept of object permanence (Baird et al., 2002). Between 12 and 18 months, children's behavior starts to incorporate tertiary circular reactions, where they will now both take greater interest in and experiment with novel objects. Finally, before the end of the 24th month, it is expected that children would have developed some form of mental representation and symbolic thought, whereby they will now engage in both imitation and make-believe behaviors.

In Piaget's theory, these different substages represent progressive incremental steps in the cognitive development of the infant (Piaget, 1952). However, it should be noted that, in other works, some of the different cognitive developments, such as the understanding of object permanence, have been found to occur in different life stages (Baillargeon et al., 1985; Kagan and Herschkowitz, 2006) and that, even within the Piagetian tradition, the very notion of progressive developmental stages has been questioned in favor of other aspects of development, such as domain specificity (Karmiloff-Smith, 1992). This would suggest that the complexity of human development implies that either the different milestones overlap or are not necessarily entirely dependent on their "predecessor," with environmental conditions also constituting an important factor influencing the outcome and process of development (Baillargeon, 1993). Alternatively, it cannot be ruled out that flaws or oversights in experimental models may have lead to different outcomes (Munakata, 2000). In any case, the milestones put forward by Piaget do seem to represent critical cognitive developments which are likely to set the foundation and facilitate the emergence of more advanced functions (Piaget, 1952; Baird et al., 2002; Fuster, 2002; Bahrick et al., 2004). It is likely then that, if the abilities gained during these stages are indeed important in the development of human cognition, then they would also be significant in the development of adaptive autonomous robots undergoing "similar" sensorimotor experiences as those related to the development of those skills (Asada et al., 2009; Cangelosi et al., 2015). This potential has led to the design of a range of different models using sensorimotor developmental principles.

In this paper, we examine the role that exposure to sensory stimuli may have on the cognitive development of an autonomous robot. Unlike related studies such as (Shaw et al., 2012; Ugur et al., 2015), which explicitly model the developmental process, the model used in our experiments was not designed following a particular sensorimotor developmental theory,

but based on a plausible epigenetic[1] mechanism (Lones and Cañamero, 2013; Lones et al., 2014). However, similar to the work of Cangelosi et al. (2015) and Ugur et al. (2015), our model leads to the emergence of an open-ended learning process achieved by allowing a robot to be able to identify and learn about interesting phenomena, a common goal of developmental models (Marshall et al., 2004). The underlying approach that our model takes to achieve the desired open-ended development also has similarities to other developmental models. Here, we use a novelty-driven approach regulated by intrinsic motivation (see Section 2.2.3). Using novelty as a way to regulate interactions with the environment and drive development has been previously explored for example by Blanchard and Cañamero (2006) and Oudeyer and Smith (2014). While there are significant differences in the way in which curiosity is generated by these models (here through hormone modulation with regard to the robot's internal and external environment), similar to Oudeyer and Smith (2014), we use the concept of curiosity to model the robot's novelty-seeking behavior by encouraging it to reduce uncertainty in an appropriate manner given its current internal state. This mechanism drives the robot's interactions, permitting it to learn and develop in an appropriate manner as it is exposed to different sensorimotor stimuli as a result of its interactions.

Using this model in an autonomous robot, we observed a natural and unforeseen developmental process somewhat similar to the sensorimotor development suggested by Piaget (1952), as we will present in this paper. As we will see, the robot's progression through, as well as the emergence of, the behaviors associated with the different developmental stages, depend on the robot's environment, and more specifically on the sensory stimuli that the robot is exposed to over the course of its development. For example, a robot placed in an environment deprived of sensory stimulation did not develop behaviors or abilities associated with the sensorimotor developmental theory. By contrast, a robot given free range in a novel environment, showed the emergence of different stages and abilities associated with the developmental process, i.e., primary and later secondary circular reactions as a consequence of its interactions in the environment. In our model, this would suggest that the emergence of these stages is related to exposure of the robot to environmental stimuli. More importantly, this leads to the research question of whether the emergence of these similar processes and stages would have any consequence for the cognitive development of the robot, or whether the similarities are simply related to a temporal phenomenon. In order to investigate this question, we allowed three robots to develop under different environmental conditions, two of which provided different levels of novelty and sensory stimulation, and the final was equivalent to sensory deprivation. We tested the robots' cognitive abilities in a range of scenarios ranging from a simple learning task to a more specific violation of expectation paradigm.

---

[1] The term "epigenetic" is used here to refer to mechanisms that lead to changes in gene expression (Schlichting and Pigliucci, 1998; Holliday, 2006; Lones and Cañamero, 2013) rather than the Piagetian notion, widely adopted in the field of developmental robotics, referring to developmental processes not directly stemming from the action of genes (Cangelosi and Riga, 2006).

The model used for these tests is based on an earlier hormone-driven epigenetic model presented in (Lones and Cañamero, 2013). This previous model showed the ability to both rapidly adapt to a range of different dynamic environmental conditions and react appropriately to unexpected stimuli. However, this early epigenetic model was based on reactive behaviors and lacked cognitive development, limiting the ability of the robot to produce and engage in planned behaviors, or to explicitly learn about new aspects of its surroundings; thus, the robot was partly dependent on information about the environment "pre-coded" in its architecture. In an attempt to overcome these limitations, we integrated in the robot's architecture a new form of neural network that we have since called an "Emergent Neural Network" (ENN), in which nodes and synaptic connections between them are created as the robot is exposed to stimulation. This network should therefore allow the robot to learn about different aspects of the environment with regard to the affordances they provide (Lones et al., 2014).

This extended model uses the same hormonal system as we used previously in (Lones and Cañamero, 2013), this time to modulate the development of the ENN in an "appropriate" manner dependent on the interaction with the external environment. The proposed model not only allowed the robot to learn about different aspects of its environment and engage in planned behavior in an appropriate manner but also gave it the ability to shape its body representation due to its interaction with the external environment. This allowed the robot to adapt successfully to a range of simple, but "real-world" and dynamic environments such as our office environment (Lones et al., 2014). However, the ENN used in (Lones et al., 2014) while successful, had constraints imposed on it due to the focus of that particular study, where the interest lay in the roles that homeostasis and hormones played on modulating curiosity and novelty-seeking behavior. Specifically, the network was explicitly pre-trained in a sterile environment and then frozen, removing the potential for additional learning. In contrast, here, where the interest lies in the roles that stimulation has on the robots' "cognitive development," the network did not undergo any pre-training and was fully active throughout. This means that the robot's learning is dependent on its own sensorimotor experiences within its environment. As we will demonstrate, for this particular model, the quality of these sensorimotor experiences is paramount to the robots' cognitive development, where a robot which has been exposed to rich sensorimotor experience develops not only greater cognitive abilities but also goes through developmental stages that we had not anticipated and which bear some similarities to infant development.

2. MATERIALS AND METHODS

2.1. Robot and Sensors

For the experiments reported in this paper, we used the medium-sized wheeled Koala II robot by K-Team. This robot is equipped with 14 infra-red (IR) sensors spread around the body, which are used to detect the distance, size, and shape of different objects. In lieu of traditional touch sensors, the IR sensors were also used to detect contact. In addition, to complement these sensors, a Microsoft LifeCam provided a microphone for sound

detection and, along with the OpenCV library, simple color-based vision. The robot's architecture was written in C++, and control of the model was handled through a serial connection to a computer running Ubuntu.

2.2. Architecture of the Robot

The software architecture giving rise to the behavior of the robot combines three main elements: a number of survival-related homeostatically controlled variables that provide the robot with internal needs to generate behavior; a hormone-driven epigenetic mechanism that controls the development of the robot; and a novel neural network that we have named a "emergent neural network", which provides the robot with learning capabilities. These three elements of the model interact in cycles or action loops of 62.5 ms in order to allow the robot to develop and adapt to its current environmental conditions as shown in **Figure 1**. This development occurs in the following manner:

- The levels of the survival-related homeostatically controlled variables change as a function of the actions and interactions with its environment (see Section 2.2.1).
- Deficits of the different homeostatic variables trigger the secretion of different artificial hormones (see Section 2.2.2).
- The different hormones, once secreted modulate both the epigenetic mechanism and emergent neural network (see Section 2.2.3).

2.2.1. Homeostatic Variables

Homeostatic imbalances have often been linked to the generation of drives and motivation in biological organisms, providing them with a potential short-term adaptation tool (Berridge, 2004). In a similar manner, in robotics, artificial homeostatic variables have been used as an effective way of generating motivations and behaviors which can be used for adaptive robotic controllers (Cañamero, 1997; Breazeal, 1998; Cañamero et al., 2002; Di Paolo, 2003; Cos et al., 2010). We have endowed our robot with three survival-related variables that it must maintain within appropriate ranges in order to survive: health, energy, and internal temperature. The three variables decrease as a function of the robot's actions and interaction within its environment. Health is a simulated variable which decreases proportionally in relation to physical contact, as shown in formula 1:

$$\Delta Health = \begin{cases} -C & \text{if } C \geq 5 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where C is the intensity of any contact, and the value 5 represents a threshold/resistance to damage[2] that must be surpassed in order for health loss to occur. Health deficits can be recovered through the consumption of specific resources found in the environment.

Energy is linked directly to the robot's battery and decreases at an average of around 15 mAh/min, although the exact amount

---

[2]Value of less than 5 are roughly equivalent to gentle strokes or minor contact resulting in no health loss.
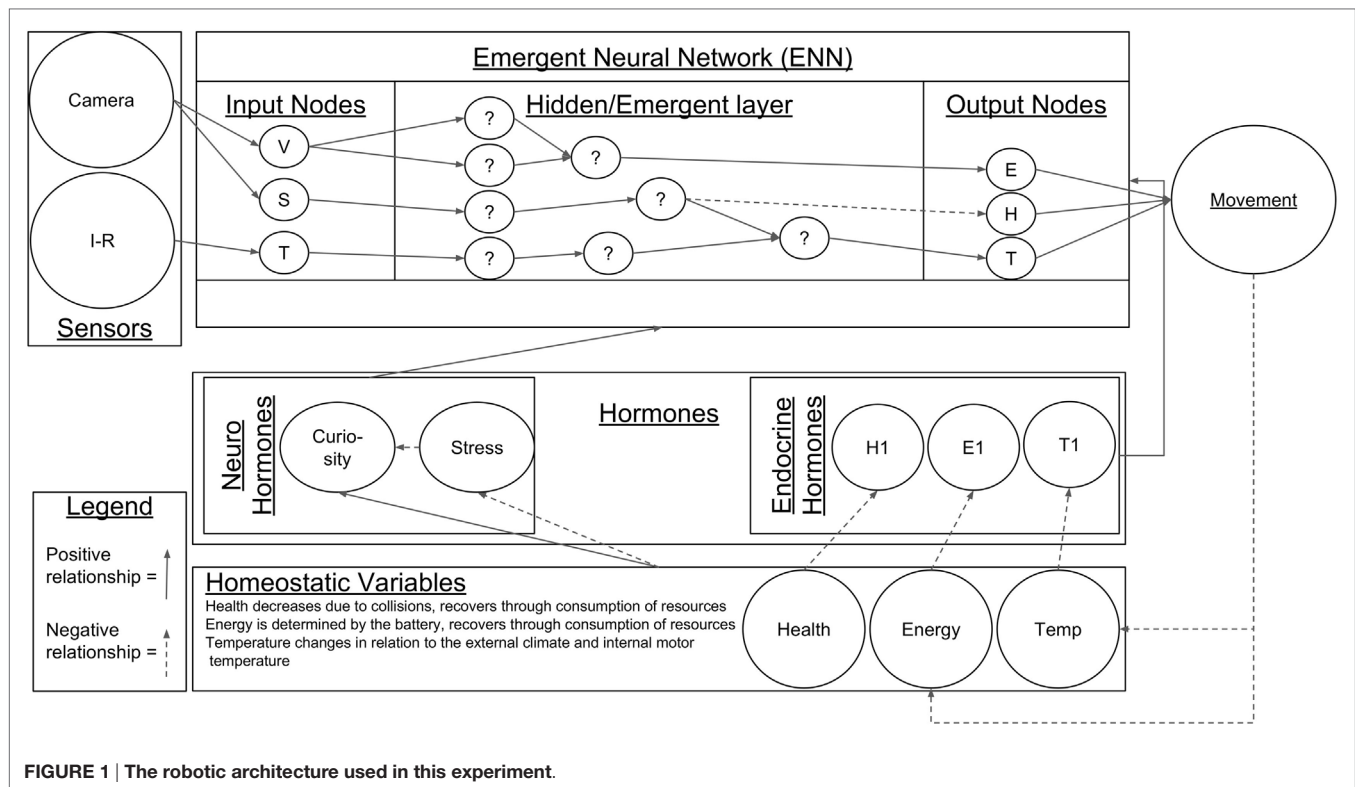
FIGURE 1 | The robotic architecture used in this experiment.

varies as a function of the robot's motor usage. Although the total battery size is approximately 3500 mAh, the robot has been programed to only sense a maximum charge of up to 75 mAh (around 5 min of running time) at any given time, creating a virtual battery. This allows us to implement a virtual charging system where the robot needs to find specific energy resources in order to recharge its virtual battery back to the maximum 75 mAh.

For both the energy and health resources, the robot recovers 7.5 and 10 U (10% of maximum capacity), respectively, per action loop when the resource is directly in front of the robot and within roughly 10 cm distance. Finally, the robot's internal temperature is related to the speed of the motors and the climate, following equation (2):

$$\Delta Temperature = \frac{|speed|}{10} * Climate \qquad (2)$$

where |speed| is the current absolute value of speed of the wheels (measured in rotations per action loop) and 10 is a predetermined constant to regulate the temperature gain with regard to movement. Climate refers to the external temperature, usually measured using a heat sensor; however, for these experiments, in order to remove unwanted variations, this was set to be detected as a constant of 24 [i.e., 24°C (75°F)].

The robot's body temperature is set to dissipate at a constant rate of 5% of the total internal temperature per action loop. Dissipation is the only method available to the robot to reduce its body temperature, meaning that in order to cool down, the robot must either reduce or suppress movement.

Each of the survival-related homeostatic variables has a lethal boundary which, if transgressed, results in the robot's death. In the case of energy and health, the lethal boundary is set at the bottom end of the range of permissible values (0), in the case of temperature the lethal boundary is at the upper end of the range (100).

2.2.2. Hormone-Driven Epigenetic Mechanism

While survival-related homeostatic imbalances, as previously mentioned, are often used to model motivation and drive behavior in autonomous robots, these imbalances alone may not be enough to ensure adaptive behavior in dynamic or complex environments (Avila-Garcia and Cañamero, 2005). A suggested, and so far successful, addition to the previous architecture, consists of integrating different hormone or endocrine systems into the model (Avila-Garcia and Cañamero, 2005; Timmis et al., 2010; Lones and Cañamero, 2013). These systems borrow from biological examples, where neuromodulatory systems have been shown to regulate behavior and allow rapid and appropriate responses to environmental events (Krichmar, 2008). These models have been used for a range of environments and conditions such as single-robot open field experiments (Krichmar, 2013) to multiple robot setups such as predator–prey scenarios (Avila-Garcia and Cañamero, 2005) to robotic foraging swarms (Timmis et al., 2010).

In our setup, we use a model consisting of five different artificial hormones, both "endocrine" (eH) and "neuro-hormones" (nH) to help the robot maintain the three previously discussed homeostatic variables. While both hormone groups share

common characteristics, they also present some significant differences. The first group of hormones (eH) is secreted by glands as a function of homeostatic deficits; each of the three homeostatic variables has an associated hormone. These hormones – E1 associated with energy, H1 associated with health, and T1 associated with internal temperature – are secreted as shown in equation (3):

$$eHSecretion_h = \psi_h Deficit_h \qquad (3)$$

where $eHSecretion_h$ is the amount of hormone h secreted, $\psi_h > 0$ is a constant regulating the amount of hormone h secreted by the gland (it can be thought of as reflecting the gland's "activity level"), and $Deficit_h$ ($0 \leq Deficit_h < 100$) is the value of the relevant homeostatic variable's deficit.

These hormones play a key role, as the robot is unable to directly detect the values of own homeostatic deficits; rather, the concentrations of the different hormones are used to signal homeostatic deficits through modulation of the ENN as discussed in Section 2.2.3, leading to the generation of drives and motivations.

The second group of hormones (nH) consists of two hormones: curiosity (nH_c) and stress (nH_s). These two hormones, which are secreted in relation to internal and external stimuli, are loosely based on the hormones Testosterone and Cortisol. nH_c will encourage outgoing behavior by increasing novelty seeking and suppressing detection of perceived negative stimuli. As an example, common behaviors linked to a high concentration of this hormone would be interest and interactions with novel objects. In contrast, nH_s will reduce novelty-seeking behavior and heighten the detection of any negative stimuli. An example of a behavior linked to a high concentration of this hormone would be the withdrawal to a perceived area of safety, which in our experimental setup often consisted of the edge of the environment – a wall – leading to the emergence of a sort of wall-following behavior. The robot's perception of walls providing safety arises here due to their perceived lack of novelty and they offer protection to one side of the robot. For the full implementation of the nH hormone group, please see Lones et al. (2014), we briefly summarize the behaviors of the hormones in equations (4) and (5).

$$nHSecretion_c = \frac{pS + \sum_v r_v}{nH_s} \qquad (4)$$

where pS is the sum of all perceived "positive" stimuli, $r_v \geq 0$ is the (perceived) recovery of a homeostatic variable v during the current action loop, and nH_s is the concentration of the stress hormone which suppresses the secretion of nH_c. By "positive stimuli," we refer to the stimulation associated (by the robot's neural network) with the recovery (i.e., the correction of the deficit) of a homeostatic variable. In other words, positive stimulation pS is the sum of any output associated with the recovery of a homeostatic variable and is calculated by the synaptic function of the output nodes of the neural network, as shown in equation (12).

$$nHSecretion_s = roD \times oS \times nS \qquad (5)$$

where roD or the "perceived risk of death" is the sum of all homeostatic deficits, oS or "overall stimulation" is the sum of the total

amount of stimulation (regardless of its type), and nS is the sum of perceived "negative" stimuli. By "negative stimuli," we refer to the stimulation associated (by the robot's neural network) with the worsening (i.e., the increase of the deficit) of a homeostatic variable. In other words, negative stimulation nS is the sum of any output associated with the worsening of a homeostatic variable and is calculated by the synaptic function of the output nodes of the neural network, as shown in equation (12). The overall stimulation oS is also determined by the synaptic function of the output nodes of the network and is the sum of the total synaptic output.

Once secreted, all hormones decay at a constant rate shown in equation (6).

$$hC_{h,t+1} = 0.95 \times hC_{h,t} \qquad (6)$$

where $hC_{h,t+1}$ is the hormone concentration in the next action loop.

The second aspect of the hormonal system is the inclusion of 5 hormone receptors, each one associated with a specific hormone. These receptors are part of the ENN and detect the current concentration of their relevant hormone (see Sections 2.3.2 and 2.3.3). The sensitivity of these hormone receptors is not constant; rather, it is modulated by a hormone-driven epigenetic mechanism (Crews, 2008; Zhang and Ho, 2011) that we introduced in Lones and Cañamero (2013). This epigenetic mechanism consists of a feedback loop where the concentrations of the different hormones will lead to either upregulation (increased sensitivity) or downregulation (decreased sensitivity) of their respective receptors, following equation (7):

$$senS_{h,t+1} = senS_{h,t} \times \frac{hC_h}{\sigma} \qquad (7)$$

where $senS_{h,t+1}$ is the hormone receptor's sensitivity in the next action loop, hC_h is the relevant hormone concentration, and σ a constant value that regulates the speed of the epigenetic change.

2.2.3. Emergent Neural Network

The emergent neural network consists of a novel design in which nodes are created as a function of the robot's interactions and exposures to different environmental stimuli. This emergent neural network, of which an example can be seen in **Figure 2**, is designed to allow the robot to learn the affordance of different aspects of its environment. Here, the term "affordance" is used in the context of the robot learning the potentialities of an action or interaction with different aspects of its environment in relation to its current internal state. Since the internal state of the robot presented here is dependent on and made up of the three homeostatic variables (see Section 2.2.1), the affordances learned by the robot will be in relation to the ability of actions to affect these said variables. For example, a potential action involving the energy resource will likely have an affordance associated with energy recovery. At this stage, it is important to highlight two aspects of the robotic model:

- First, all behaviors discussed here emerge as a result of the development and modulation of the neural network, simply put there are no pre-designed behaviors or internal states.
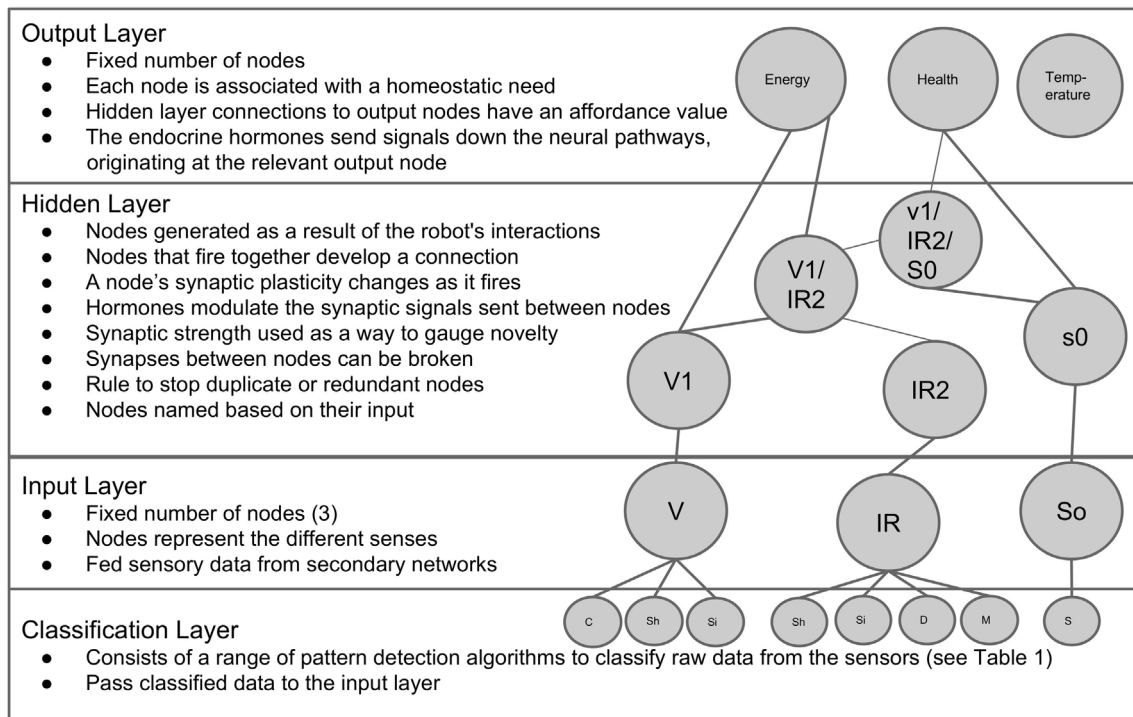
FIGURE 2 | An example of the basic structure of the neural network.

- Second, the development of both the neural network and affordances are based on the robot's perceptions and interactions, therefore robots with different morphological designs or placed in different environmental conditions will likely develop in different ways.

2.3. Detailed Description of the Emergent Neural Network

The emergent neural network used in this paper consists of a three-layer network design shown in **Figure 2**. The first layer of the network consists of an input layer which is fed sensory data from a range of different classification networks. The second layer is the hidden layer in which nodes emerge as a function of the robot's interactions and environmental exposures. This layer is responsible for recognizing different aspects of the environment and assigning an appropriate affordance based on the robot's past experiences. The final layer is the output layer which simply sums the detected affordances.

2.3.1. Classification and Input Nodes

The input layer consists of three fixed nodes, each representing one of the robot's different sensory modalities. These modalities are vision, IR, and sound and receive input data from different pre-processing classification algorithms shown in **Table 1** and **Figure 3**. These three *input nodes* are quite different to conventional neurons found in other networks. In our network, these nodes will fire differently depending on which classification

network is currently feeding input, with each node in the input layer associated with a specific fixed group of classification networks (see **Table 1**). For example, the node representing the vision modality is associated with classification networks that detect color, shape, and size. These input layer nodes work as follows.

For each sensory modality, the output from each of the pre-processing classification networks (shown in **Table 1**) consists of a 4-digit input pattern that feeds into the appropriate node in the input layer. The four digits provide information about the sensory modality used, the type of stimulus, the position of the stimulus with respect to the body coordinates of the robot, and the number of instances of that stimulus detected in that action loop. The number of pre-processing classification networks associated with each node of the input layer depends on the modality of the latter – three for vision, four for IR, and one for sound (see **Table 1**). For each input pattern received, each node in the input layer will either strengthen its connection with a node in the hidden layer corresponding to that input pattern if a node has already been associated with it, or create a new node if the pattern is classified as novel. In any one time frame, a node in the input layer can receive multiple inputs from each pre-processing classification network, and thus it can potentially create multiple new nodes in the hidden layer.

As an example, when perceiving the face depicted in **Figure 4**, the vision node in the input layer would receive an input from the shape pre-processing classifier consisting of the four digits 1 (indicating the vision modality), 0 (representing a circle),

TABLE 1 | The different sensory modalities and their implementation and design.

	Vision			IR				Sound
Sense	Color	Shape	Size	Shape	Size	Distance	Movement	Volume
Sensor	Webcam	Webcam	Webcam	IR	IR	IR	IR	Webcam
Algorithm	OpenCV HSV	OpenCV contour	OpenCV contour	Pattern detection	Pattern detection	IR value	Compare IR value	Sound
Values	White = 0	Circle = 0	Small = 0	Flat = 0	Small = 0	Close = 0	None = 0	Silent = 0
	Black = 1	Square = 1	Medium = 1	Curved = 1	Medium = 1	Medium = 1	Small = 1	Quiet = 1
	Green = 2	Rectangle = 2	Large = 2	Corner = 2	Large = 2	Far = 2	Medium = 2	Medium = 2
	Red = 3	Triangle = 3	Unknown = 9	Hole = 3	Unknown = 9	Unknown = 9	Large = 4	Loud = 3
	Yellow = 4	Crescent = 4		Unknown = 9			Unknown = 9	
	Blue = 5	Unknown = 9						
	Unknown = 9							
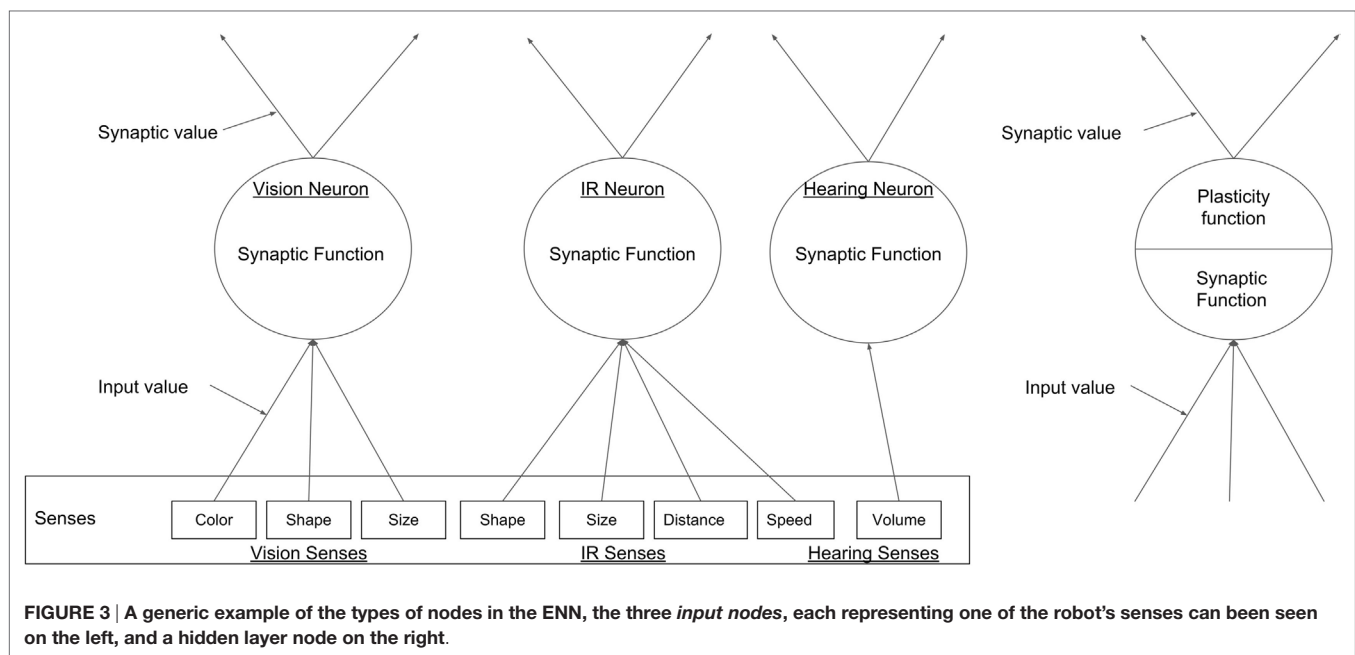
FIGURE 3 | A generic example of the types of nodes in the ENN, the three *input nodes*, each representing one of the robot's senses can been seen on the left, and a hidden layer node on the right.
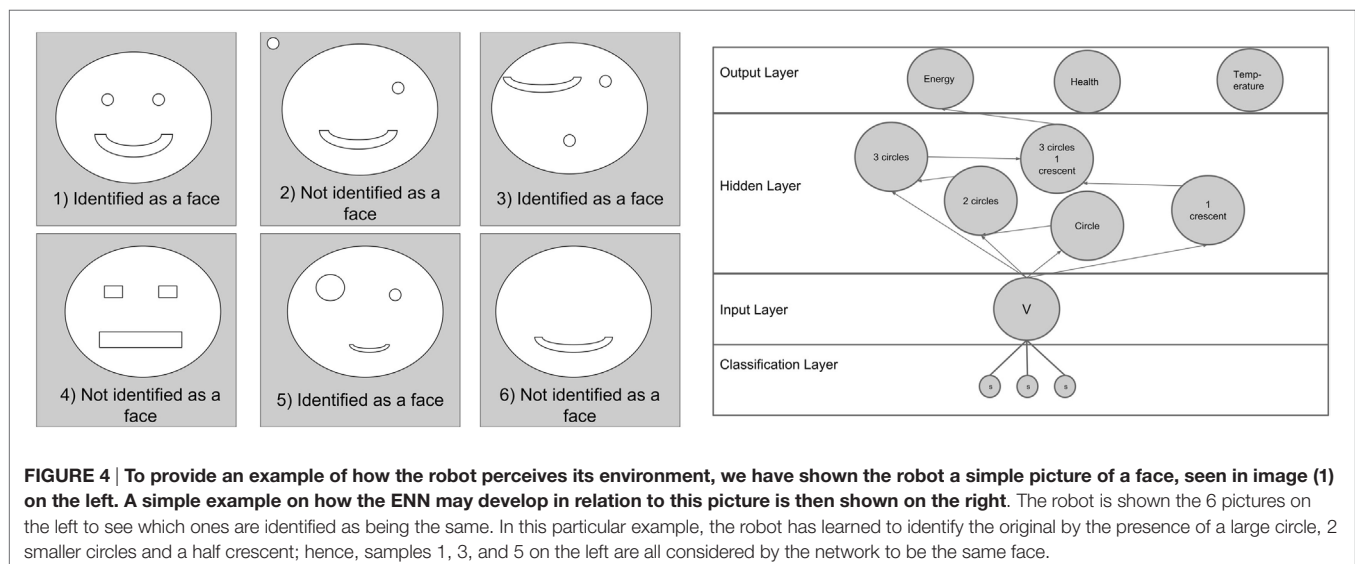
FIGURE 4 | To provide an example of how the robot perceives its environment, we have shown the robot a simple picture of a face, seen in image (1) on the left. A simple example on how the ENN may develop in relation to this picture is then shown on the right. The robot is shown the 6 pictures on the left to see which ones are identified as being the same. In this particular example, the robot has learned to identify the original by the presence of a large circle, 2 smaller circles and a half crescent; hence, samples 1, 3, and 5 on the left are all considered by the network to be the same face.

4 (if the face was directly ahead), and 3 (for the three circles: two eye circles plus the larger enclosing circle). It would additionally receive an input of 1, 4, 4, and 1 (indicating, respectively, vision, crescent, ahead, and one instance).

2.3.2. Hidden Layer

The second layer of the ENN is the hidden layer, which receives data from the input layer and sends data to the output layer. This layer initially starts empty, and nodes are created as a function of the robot's exposure to different stimuli. Creation of nodes takes place under two circumstances:

1. When an input node fires but has no synaptic connection to a relevant node (as described in the section above), or
2. When two or more hidden layer nodes fire at the same time.

When a new node is created in the hidden layer, in addition to being connected to the relevant nodes that led to its creation (which provide the input), it is also fully connected to the nodes of the output layer. However, all these different connections can disappear as the network continues to develop. When a synaptic connection between two nodes i and j is created, it is given a strength of $sP_{ij} = 0.5$. The connection strength is then updated as the robot interacts with its environment, using a sigmoid function, as seen in equation (8).

$$sP_{ij} = \alpha e^{\beta e^{x_{ij}e}} \tag{8}$$

where α and β are constants, and x_{ij} is the sum of times the nodes i and j have fired together, minus the number of times that they have not fired together, within a range of $-10{,}000 < x_n < 10{,}000$. A negative value of x_{ij} thus means that, more often than not, the nodes have not fired together.

Equation (8) results in a synaptic connection with strength in the range $0 < sP_{ij} < 1$. Due to the sigmoid nature of the function, the closer the synaptic strength gets to either end of this range, the lower the rate of change, or plasticity, of the connection. The synaptic strength of a connection between nodes plays a number of roles in this ENN, as will be discussed shortly. One of the most important roles is simply determining if a connection exists between nodes. This is achieved as follows:

- If a synaptic connection between two nodes exists but the synaptic strength drops below 0.05, then the connection is broken.
- If a synaptic connection does not exist but synaptic strength would be above 0.1 if it existed, then a connection is made.

In addition, when a synaptic connection is made from a new node to a node in the output layer, this connection is assigned an affordance – the potential to recover a homeostatic variable. Initially, this affordance is set at the change detected in the related output node's homeostatic variable. For instance, a new synaptic connection to the energy output node during a loop when the robot gained 2 units of energy will result in that connection receiving an initial affordance value of 2. This affordance assigned to the synaptic connection then changes as the robot continues

to interact with that particular aspect of the environment, as follows:

$$\Delta Affordance_{v,i,j} = Affordance_{v,i,j} \times sP_{ij} \\ + HomeostaticChange_v \times (1 - sP_{ij}) \tag{9}$$

where $HomeostaticChange_v$ is simply the change in a homeostatic variable v in the current action loop compared with the previous one.

In order for a node i in the hidden layer to fire, it must receive a total input that is greater than or equal to its total number of synaptic inputs, thus:

$$output_{i,d} = \begin{cases} 1 & \text{if } input_{i,d} >= sC_i \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where sC_i is number of input synaptic connections for node i, and d ($0 \le d < 8$) is the direction of the detected stimulus with respect to 8 equally spaced body coordinates of the robot, the third digit of the input pattern discussed in Section 2.3.1. Using this system, 0 represents the body coordinate directly behind the robot, and then going clockwise each subsequent value represents the next coordinate. For example, 4 represents a coordinate directly in front of the robot.

As shown in equation 10, if the firing threshold is reached, the nodes fire with a value of 1; however, the synaptic function $sF_{i,j,d}$, or output of the hidden layer node i, is then modified depending on the outgoing connection of the node and the directional origin of the stimulus d. If a hidden layer node i is connected to another hidden layer node j, the synaptic function is:

$$sF_{i,d} = output_{i,d} \times nHmodulation_i \times \sum_v eHModulation_{v,i} \tag{11}$$

where $eHModulation_{v,i}$ is the modulation of the endocrine hormones eH_v on node i [see equations (13) and (14)] and $nHmodulation_i$ is the combined strength of the modulation from the neuro-hormones stress and curiosity nH [see equation (15)]. The roles of hormones in the ENN are discussed in greater detail in Section 2.3.3.

If the node is connected to an output node for homeostatic variable v then the synaptic function is given by:

$$sF_{i,j,v,d} = nHmodulation_i \times output_{i,d} \\ \times Affordance_{v,i,j} \times eHModulation_{v,i} \tag{12}$$

A basic example of how the ENN works and allows the robot to identify objects and stimuli can be seen in **Figure 4**, which shows how the robot perceives a simple face. Here, the robot is able to identify the face by the presence of the key characteristics of a large circle, 2 small circles and a crescent. However, the robot cannot detect spatial arrangements, and therefore, as long as the features are close enough, they will be identified as the same object. The characteristics used by the robot to identify objects depend on its past learning. A relatively new robot, for instance, may identify all pictures as being the same, since they all posses a circular shape. In contrast, a robot with greater environmental exposure, such as the one used in this example, will have more specific criteria.

2.3.3. Hormones in the ENN

As shown in equations (11) and (12), different hormone concentrations modulate the synaptic functions of the ENN. As discussed in Section 2.3.2, these hormones modulate the nodes within the ENN as a function of the internal state of the robot. In the case of the eH hormones, the strength of the modulation is dependent on the hormone's sensor's sensitivity and the connections of the nodes. For a node i directly connected to the output layer, the modulation by the eH_v is given by:

$$eHModulation_{v,i} = eHConcentration_v \times senS_v \quad (13)$$

where $eHConcentration_v$ is the concentration of the eH hormone eH_v, $senS_v$ is the sensitivity to eH_v [see equation (7)].

However, for nodes not directly connected to an output node, the modulation from the eH becomes weaker, as shown in equation (14), resulting in a larger modulation of the nodes closer to the output layer and/or with stronger synaptic connections to it. Using hormonal modulation in this manner promotes the activation of nodes that have a higher synaptic strength, and hence promotes behaviors that, in past interactions, have led to better homeostatic balance.

$$eHModulation_{v,i} = \sum_{j \in O(i)} \frac{eHModulation_{v,j} \times sP_{i,j}}{noI_j} \quad (14)$$

where $O(i)$ is the set of output nodes from node i, i.e., the set of nodes that are connected to output of i, $eHModulation_{vi}$ is the strength of the modulation in the current node, dependent on the sum of the signal passed down from connecting nodes $eHModulation_{vj}$, and noI_j the number of input connections of node j.

In contrast to the eH hormones, the nH hormones surround the ENN, affecting all nodes equally. The nH behave differently as their role is to either promote or suppress novelty-seeking behavior. This is caused by the combined effect of the curiosity and stress hormones. The curiosity hormone increases the activation of nodes with a low synaptic strength and suppresses nodes with a high synaptic strength. Conversely, the stress hormone increases the activation of nodes with a high synaptic strength and suppresses nodes with a low synaptic strength, as in equations (11) and (12). Therefore, the robot is using the synaptic strength as a way of assessing the novelty value of an object or aspect of the environment, since a high synaptic strength only happens if an object behaves as expected each time the robot interacts with it. This can be seen below in equation (15)

$$nHmodulation_{ij} = sP_{ij} \times nHConcentration_s \times senS_s \\ + (1 - sP_{ij}) \times nHConcentration_c \times senS_c \quad (15)$$

where $nHConcentration_s$ is the concentration of the stress hormone (s), $senS_s$ is the receptor's sensitivity to the stress hormone, $nHConcentration_c$ is the concentration of the curiosity hormone (c) and $senS_c$ the receptor's sensitivity to the curiosity hormone.

2.3.4. Output Layer

The final layer of the ENN is the output layer, which consist of a fixed number of nodes equal to the total number of survival-related homeostatic needs. Each output node simply sums up the total input from the hidden layer in order to calculate the affordance of moving in a certain direction.

$$output_{v,d} = \sum_i sF_{i,v,d} \quad (16)$$

The output of the ENN then feeds directly into (and modulates) the robot's actuators – in this case, the wheels:

$$WheelSpeed_i = \sum_{v,d} output_{v,d} \times set_{i,d} \quad (17)$$

where $WheelSpeed_i$ is the speed of the left (0) or right wheel (1), set_{id} are constant vectors equal to $(-10, -10, -5, -3, 1, 3, 5, 10)$ if $i = 0$, or $(-10, 10, 5, 3, 1, -3, -5, -10)$ if $i = 1$. This means that if a single stimulus originating from the left side of the robot is detected, the robot's left wheel moves at a speed of $5 \times output$ and the right wheel moves at a speed of $5 \times output$. Therefore, a positive output will result in the robot turning toward the stimulus and a negative output in turning away from it.

To summarize, the causal chain that leads to internal or external stimuli promoting different behaviors is as follows:

1. As homeostatic deficits occur they lead to the release of the associated endocrine hormone eH [see equation (3)].
2. Internal and external stimuli lead to the release of the neurohormones, with curiosity being secreted in relation to perceived positive stimulation and stress in relation to negative stimulation [see equations (4) and (5)].
3. The robot's sensitivity to each hormone is dependent on its historic exposure to it [see equation (7)].
4. The artificial hormones modulate the synaptic function of the hidden layer nodes [see equation (12)] depending on the nodes position in the network [see equations (13) and (14)].
5. The output nodes sum up the synaptic function of connected neural pathway [see equation (16)], with the value dependent on past outcomes associated with pathways activation [see equations (8) and (9)].
6. The output from the network then affects the behavior (wheel speed), by promoting or suppressing the tendency to move in a certain direction at a certain speed. The larger the output, the greater modulating effect it will have on behavior [see equation (17)].

2.4. Experimental Setup

To test the effects of the previously described robot architecture, we allowed the robot to develop in three different environments with a single run of 60 min of duration in each. These three environments were (1) a base/standard environment, (2) a novel environment, and (3) a sensory deprivation environment. For each of these environments, the robot spent the first 10 minutes with a caregiver who looked after it, and introduced the robot to core components of the environment. In the remaining 50 minutes, the robots were then placed in their specific environments. The base and novel environments both consisted of our open lab environment (see **Figure 5**) with some differences that we will discuss in the relevant sections (Sections 1 and 2). In the sensory deprivation environment, the robot was placed and left

FIGURE 5 | Different aspects of the environment used during the experiments. (1) shows a panoramic picture of the standard open environment used during 3.1 and 3.2. (2) shows an example of one of the novel structures used during experiment 3.2. (3) shows the koala robot used during this work and the cardboard box the robot was placed in to create a sensory deprivation environment. (4) shows two AIBO robots used as novel objects in the test described in Section 4.2.

alone inside a cardboard box after the initial 10 minutes with the caregiver. In the three environments, the robot would have access to two sources of each type of resource; in the third experiment, this meant that the resources were placed inside the box along with the robot.

3. EXPERIMENTS AND RESULTS

Before the robots were placed into their different environments, each spent the first 10 min of their "life" with the caregiver. This caregiver provided an identical experience for each of the robots with the primary purpose to teach them the critical aspects needed to survive, such as how to recover from homeostatic deficits. This period essentially consisted of the caregiver sating the robot's needs by bringing the relevant resource to them. During this period, the robot's behaviors were essentially driven by exposure to stimuli. In some ways, these basic behaviors are similar to the so-called reflex acts of a newborn. At this stage, both newborns

and the robots display many "reflex" behaviors; for instance, a newborn will "grasp" objects placed into their hand or suck an object placed against their lips; our robots' "reflexive" behaviors will generally see them move toward or away from (attraction vs. repulsion) different environmental stimuli.

In this first phase, the interactions between the robots and the caregiver resulted in the emergence of five main reflex behaviors. The first three occur due to the homeostatic variables, which are Attraction/Repulsion, Avoidance, and Recoil. Attraction and Repulsion emerged when the caregiver fed the robots by placing a relevant resource in front of them. This "feeding" by the caregiver, made the robot move toward the caregiver when hungry and then away when sated. Avoidance emerged when the caregiver moved too close to the robot, making the robot move toward an area with more space. Recoil emerged when physical contact occurred; unlike with the avoidance behavior, here the robot will prefer to move in an opposite direction to the stimulus rather than simply toward more space.

The final two reflex behaviors seen during this period are slightly different. The Exploration behavior emerges due to a combination of the first three behaviors: the Attraction behavior gives the robot the motivation to move forward while the Avoidance and Recoil lead to a motivation to avoid collisions. Finally, Localized Attention, the last innate behavior seen during this period, is based partly on learning and emerges around the 8-minutes mark. This behavior sees the robot turn to face a moving object that is roughly within a 30-cm range. The basis behind this behavior can be traced to the fact that the robot at this stage associates movement with the presence of the caregiver[3] and therefore the impending "feeding," which can only occur if the robot is facing the resource (and hence the caregiver holding it). At the end of this initial period, the caregiver would leave the environment and be outside the robot's view.

3.1. First Experiment: The Standard Environment

In the first of the three experiments, the robot was placed in our open lab environment shown in **Figure 5**. For this experiment, the robot was given free rein of our lab with only limited changes to the environment made. These changes include (a) the use of plywood borders to block access to "problem" areas where the robot's sensors and actuators would be unsuitable and (b) the placement of resources. Additionally, blackout curtains were used to block natural light, in order to keep lighting conditions comparable through the experiments.

3.1.1. First Experiment: Minutes 10–20

During this period of the experiment, the caregiver was removed from the environment, and with him the feeding interaction between the caregiver and the robot. From now on, in order to maintain its homeostatic balance, the robot would need to seek out the different resources scattered throughout the environment. Resources were placed in manner in which they could be clearly seen by the robot – four resources, one in each corner, alternating in type – with the aim of causing it to move around the environment in order to experience different sensorimotor stimuli. The prior 10-min exposure to resources through the caregiver's feeding was enough for the robot to have begun to learn some of the key features of the different resources, such as their shape, color, and size, to allow the robot to detect them.

The immediate challenges that the removal of the caregiver presents to the robot are threefold. First, the robot must be able to manage conflicting needs, e.g., if it chooses to replenish energy it must at least temporarily forgo reducing its temperature or replenishing its health. Second, the robot needs to develop tolerance so its consumption pattern – particularly to what level it can let a homeostatic variable drop before replenishing – is appropriate for the current environment. Third, the robot must adapt its sensorimotor behavior – how fast to move and when to turn to avoid collisions – to the current environmental conditions.

At the beginning of this period, the robot was highly sensitive to its internal needs – attempting to replenish any variable that was roughly below 90%. Due to spacing of the resources, the robot was often able to see at least one of each type at any given time, and therefore at this point, it did not search the environment when a deficit occurred but rather moved to the nearest perceived resource. This movement was often inefficient (see **Figure 6**) as in many cases a closer resource was located outside its immediate field of view, either to the side or behind. However, at this point in time, the robot's behavior was still largely reflex-driven – seeing the resource made the robot move toward it. When two homeostatic variables were low and the required resources could both be seen, the robot's choice of which variable to recover first would be determined based on the size of both the internal deficit and the detected stimuli. A problem with satisfying needs in this manner is due to a combination of noise – the perceived size the of external stimuli would fluctuate – and homeostatic variables not decreasing linearly or at an equal rate; the robot's intrinsic motivations would thus fluctuate, and hence its "goals" and executed behaviors often changed before a need was satiated as shown in **Figure 7**.

3.1.2. First Experiment: Minutes 20–30

The inefficiency in the robot's behavior after the withdrawal of the caregiver initially leads to the robot having issues in maintaining homeostasis. However, after the robot had been sufficiently exposed to its environment and the epigenetic mechanism began to regulate hormone receptors, its behaviors became more appropriate, and the robot was able to recover a homeostatic deficit 54% faster on average. This can been seen in **Figures 6** and **7** which show, respectively, the change in the robot's movement patterns and motivations.

As shown in **Figures 6** and **7**, the robot's movements have become much more efficient for its environment, as it now moves more directly between the resources with limited motivation or behavior switching. This occurred first as a result of a change in tolerances to homeostatic deficits. As the robot had consistently lower but stable homeostatic variables due to needing to feed for itself, it soon became tolerant to these lower levels through the epigenetic mechanism. This resulted in reduced urgency in replenishing its internal variables, to the extent that they would now need to reach an average level of around 60% instead of the previous 90% before the robot would become motivated to replenish them. As a consequence of the reduced need to replenish the homeostatic variables of energy and health, the robot was no longer under such internal pressure to move quickly between the resources and could reduce its overall speed, resolving the issues of overheating and increased collisions associated with faster movement in the previous period. Additionally, while the robot maintained a relatively constant speed in previous periods, slowing down only to consume or due to internal overheating, now the robot began to modulate its speed to match the environmental conditions. For example, the robot would move slower near the edges of the environment where it previously had collisions, and faster in the open middle areas.

This period represented an important time in the robot's development. As described previously, during the early stages of

---

[3] Although the robot did posses color vision at this stage, perhaps due to environmental noise or to slower development of vision, the robot relied on movement to detect the caregiver.
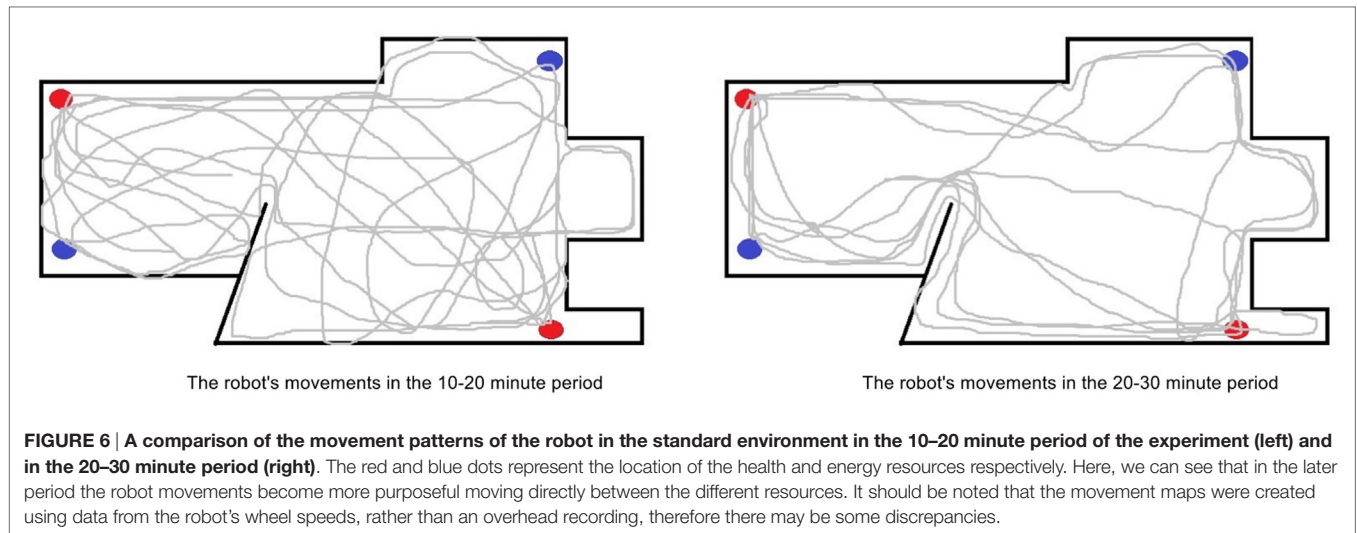
FIGURE 6 | A comparison of the movement patterns of the robot in the standard environment in the 10–20 minute period of the experiment (left) and in the 20–30 minute period (right). The red and blue dots represent the location of the health and energy resources respectively. Here, we can see that in the later period the robot movements become more purposeful moving directly between the different resources. It should be noted that the movement maps were created using data from the robot's wheel speeds, rather than an overhead recording, therefore there may be some discrepancies.

this experiment, when the robot was first exposed to this environment, its behavior was almost entirely reflex-driven. However, due to motor stimulation, the robot's behavior has started to become adaptive, taking into account the current environmental conditions and its own physical body.

This period therefore potentially bears some similarities to the concept of primary circular reaction in infant development. Much like with infants at this stage, here the robot's focus is on the effects that its behaviors had on its own body – for instance, developing appropriate movement speeds and understanding and adapting to the restraints of the levels of its homeostatic variables. Similarly, for both the robot and the infant, behaviors categorized as primary circular reactions emerge as accidental discoveries (Papalia et al., 1992; Schaffer, 1996).

3.1.3. First Experiment: Minutes 30–40

During the first 30 minutes or so, the robot had begun to adapt its behaviors with regard to maintaining homeostasis by developing behaviors which have similarities to primary circular reactions. However, at this point, the robot began to show the emergence of more complex behaviors that could be considered similar to secondary or even tertiary circular reactions, as we will discuss in more detail below. At around the 33 minutes mark, due to the robot's previously discussed reduced need for, and increased efficiency in, maintaining homeostasis, the robot spent a much smaller proportion of its time attending to homeostatic needs, showing a reduction from 93% of its time actively searching for resources in the first 30 minutes down to 59% in this period shown in **Figure 8**. This reduction in time needed to maintain homeostasis provided the robot with the opportunity to *explore* and interact with other aspects of the environment. During this period of exploration, using the previously discussed novelty mechanism (see Section 2.2.3), the robot's motivations were determined by both the internal and external environment. Such exploration would take different forms, depending on hormonal levels. With high levels of the nH_c, which is associated with positive stimuli and a good level of

homeostatic variables, the robot's attention was focused on the novel aspects of the environment. These novel aspects tended to be objects or areas that the robot had limited knowledge of, and/or objects that had some perceived uncertainty or danger as to the outcome of any interaction. In contrast, with higher levels of the nH_s, which is associated with negative stimuli, over-stimulation and poor homeostasis maintenance, the robot is more attracted to, and will interact with, less novel aspects, such as those it already had some understanding of, or perceives to be safe, e.g., the walls of the environment due to their static nature. In cases where very high levels of the nH_s were present, the robot would simply move to an area of perceived safety and only leave when the nH_s levels had decreased sufficiently.

This period represented an important stage in development of the robot for two critical reasons. First, during this period, the increased exploration is strongly linked to the growth of the ENN (see Section 4.1). Second, this exploration and interaction represent an opportunity for the robot to further understand both its own body and the ways in which it can influence its environment. Due to the relatively static nature of this first environment – most objects were either immovable or too large for the robot to meaningfully interact with them – interaction was relatively limited; it consisted for the most part in pushing an object for a few seconds, before learning that the only outcome of this behavior was a reduction in its health due to the contact, thus reducing future attempts to interact with the said object. However, around the 38th minute, the robot found the resources which consisted of small plastic balls, light, and easy to push, and therefore the robot was able to create an interesting novel experience for itself by pushing the balls.

3.1.4. First Experiment: Minutes 40–60

During the latter stages of this experiment, due to improved efficiency in recovering homeostatic deficits, the robot spent most of the time either idle or interacting with resources. Initially, this interaction consisted of small pushes that took place over a period of around 10 minutes. The motivation for the robot to
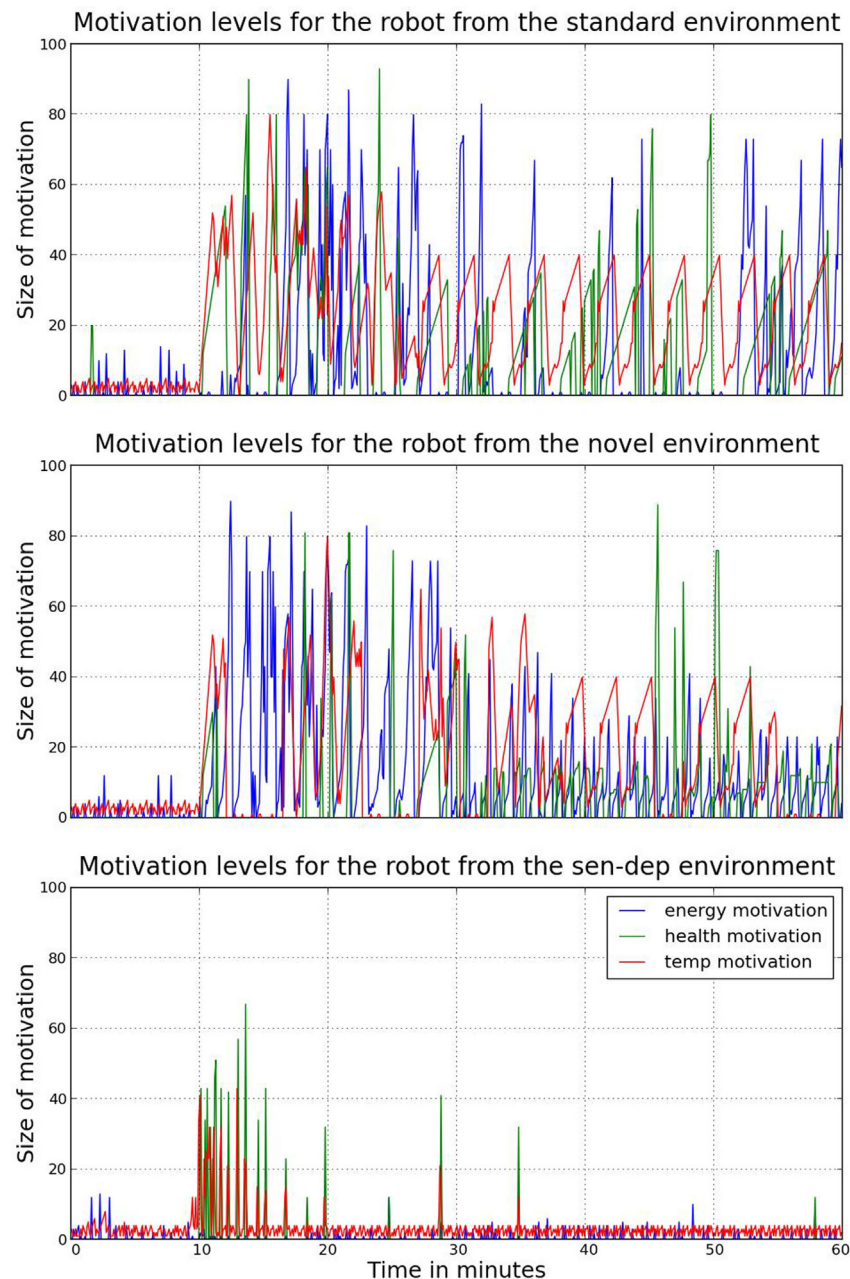
FIGURE 7 | A comparison of the three main motivations (replenishment of a homeostatic variable) for each robot during the experiments. As can be seen, during the period immediately following the initial ten minutes with the caregiver, changes and growth of motivations are much more volatile. This leads to increased occurrences of rapid behavior switching. Due to the volatility of change, this can lead to both inefficiencies and missed opportunities, i.e., constantly moving between two resources without (fully) feeding.

push the balls was twofold. Initially, the pushing was curiosity driven, as the robot tried to learn what the pushing resulted in. After around 5 minutes, however, the pushing became novelty driven, caused by the new element of motion, as mentioned in the previous section. As expected in our model, due to the high novelty that resulted from pushing an object, the robot would only "purposefully" push objects when it had high ratio of nH_c to nH_s concentration.

This emergent behavior presents some similarities with ideas of secondary circular reactions. For example, a child using a rattle and our robot pushing the ball share the fact that the agent is beginning to notice and explore that their actions and behaviors can have interesting effects on their surroundings. Similarly, later the behavior where we see the robot pushing the ball in order to create a novelty source has similarities to progression of secondary circular reaction to coordinated secondary circular reaction,

FIGURE 8 | The type of behavior executed by each robot during each 10 min period. As previously stated, the robot has no explicit behaviors; instead, behaviors executed by the robot have been classified into four general groups. Interaction includes any purposeful movement toward or contact with an aspect of the environment, foraging refers to any behavior that deals with the recovery of a homeostatic variable, this includes consuming, moving toward and searching for a resource, exploration includes any movement-based behavior, while finally inactive is any period where the robot remains stationary without consuming or engaging in interactions.

where the robot is now demonstrating the ability to manipulate an object to achieve a desired effect.

We observed another interesting phenomenon at around the 47 minutes mark, as the robot seemed to develop a search strategy while looking for resources. Previously, when searching for a resource, the robot would randomly explore its environment; however, at this point, the robot began to show some strategy in its search, since instead of the random exploration, it would now move to the walls and follow them to search for the resources, which were placed near the corners of the environment. The emergence of this behavior further reduced the average time spent searching for a resource from the previous 59% down to 47%. As time went on, this behavior continued to develop and the robot began to learn to associate certain easily identifiable

landmarks in the lab, such as a blue screen or a cupboard, with the presence of a particular resource. This ability greatly improved the time needed to find a resource, further reducing the average time spent searching for resources down to 21%. This behavior might suggest that the robot had developed some notion of "object permanence". However, it may be a simple association between resources and landmarks, which is a significantly simpler concept than object permanence. In order to investigate which of these might be the case, we carried out the experiments reported in Section 4.3.

3.2. Second Experiment: The Novel Environment

In the second experiment, we developed the robot in an environment very similar to the one used in the first experiment, with the difference of the inclusion of a range of different novelty sources. These included light movable objects arranged in various shapes and patterns, as shown in **Figure 5**, as well as two small Khepera robots that moved around randomly. If, at any time, any of these object were knocked over (e.g., due to the Koala robot's interactions) or stopped functioning as intended, the caregiver would replace or reset them as soon as the robot had moved away.

3.2.1. Second Experiment: Minutes 0–30

As we would expect, in the early stages of this experiment the exposure to additional (with respect to the first environment) sources of novelty had no real effect on the robot due to its effort to maintain homeostasis. Apart from the need to avoid the two additional randomly moving robots and the additional novel objects, the behavior and development of this robot was almost identical to the robot in the first experiment as shown in **Figures 7** and **8**. For this reason, we will not spend time discussing this robot's early life but will rather move on to the second half of the experiment, when the behavior started to deviate.

3.2.2. Second Experiment: Minutes 30–50

Much like the robot in the first experiment, at around 33 minutes into its development, this robot had adapted to its environment well enough to no longer need to spend the majority of its time looking for resources. The exception to this, shown in **Figures 7** and **8**, occurs between the 40th and 50th minutes. Due to the increased interaction with objects as discussed shortly, the robot suffers additional health damage as it learns how to properly interact; therefore, it spends additional time during this period recovering its health variable.

While the robot in the first environment spent much of its "free time" being idle simply due to a lack of things to do, (i.e., a very limited number of novelty sources to interact with), this robot had a much larger range of possible objects to learn about. As before, the robot's interest in the novel objects in the environment depended on the concentration of the nH_c and nH_s hormones. Initially, with a high value of the nH_c, the robot's attention was mostly focused on the randomly moving robots. During this period of high concentration of nH_c, in the initial instances, the robot would simply engage in a following behavior

moving behind the nearest moving robot. After around 2–3 instances of this following behavior, the robot began to intensify its interaction by engaging in both pushing and approaching the small robot from different angles. Since the randomly moving robots had been programed to stop if contact was detected, the novelty value that the robot would associate with them greatly diminished over a period of around 5 minutes, dropping to almost zero novelty near the start of the 50th minute as shown in **Figure 9**.

In contrast, with a medium concentration of the nH_c, the robot was attracted to the different arrangements of objects that were constructed with the small tin cans (see **Figure 5**). Initially, the robot would either move close to these structures or slowly circle around them. After a couple of minutes, when the robot was familiar with the structures, it began to make physical contact with them through gentle bumps and pushes. Due to the lightweight nature of the tin cans, any physical contact from the robot would easily knock them over, and this resulted

in the robot detecting not only a large amount of unexpected rapid movement around itself but also collisions, as some of the tin cans hit the robot. Since the robot only had a moderate amount of the nH_c when initially interacting with the structures, their falling resulted in significant over-stimulation, leading to increased secretion of the nH_s and the robot's withdrawal to a perceived safer location. These implications of the early contact with the structures resulted in the robot associating a higher level of perceived novelty with them due to the uncertainty of the outcome of any interaction. This increase in novelty associated with the structures along with the decrease in novelty associated with the Khepera robots resulted in structures having the highest perceived novelty as shown in **Figure 9**. Due to the increased perceived novelty, the robot would now only interact with the novel structures with high nH_c levels. The higher concentration of nH_c protected the robot from becoming overstimulated due to unpredicted outcomes, which led to more thorough interaction with the structures. In the last 5–10 minutes of this period, the robot engaged with the structures in a number of different ways as it attempted to learn about them – including moving around them at different speeds, stopping near them at different distances, trying to move through them, and pushing them with different intensities.

3.2.3. Second Experiment: Minutes 50–60

At around 54 minutes into its development, the robot started displaying a new behavior: it would gently push over a structure before moving away and stopping. As we previously mentioned, when a structure was knocked down, the caregiver would replace it when the robot had moved away. As soon as the caregiver entered the environment to replace the tin cans, the robot immediately moved toward them and tried to interact with the caregiver. The caregiver, due to a number of factors such as size, shape, and movement, was unsurprisingly perceived as highly novel by the robot (see **Figure 9**). What was, however, interesting is that the robot seemed to engage in this sequence of behaviors "on purpose." It is likely that, after experimenting with the objects, the robot had learned that by pushing the structures over, it could cause the caregiver to enter the environment and use this to satisfy its own need for novelty. Before the 54th minute, the robot had not displayed this behavior sequence of trying to have the caregiver enter the environment; yet, after the first occurrence, in the remaining 6 minutes of the experiment, this behavior occurred 11 additional times. In all cases, this behavior only occurred with high nH_c and low nH_s levels, supporting the idea that the robot was using this behavioral sequence "on purpose" to satisfy its own need for novelty. Examining the ENN seems to back up this idea, as neurons associated with the caregiver were active when interacting with the tin cans.

This behavior by the robot could be regarded as the emergence of a form of tertiary circular reactions and potentially bear a similarity to a representation of cause and effect. With regard to tertiary circular reactions, the robot was demonstrating the ability to not only manipulate and experiment with different objects in its environment, but also to use these objects in order to change

FIGURE 9 | An overview of the average perceived novelty of 5 different aspects of the robot's environment during each time period. It should be expected that as a robot interacts with an aspect the novelty value will decrease. The exception to this is if the object has unpredictable or dynamic behavior in which case the novelty value would be expected to rise as the robot interacts with it.

its environment, thus suggesting some sort of representation of cause and effect, an aspect of tertiary circular reactions (Papalia et al., 1992).

The formation of these representations is less clear, though, since the robot's behavior of knocking over structures in order to bring the unseen caregiver back into the environment could potentially suggest object permanence, which we test later in Section 4.3.

3.3. Third Experiment: Sensory Deprivation

In the final experiment, instead of being allowed to move freely in an open environment like in the previous experiments, after the first 10 minutes of interaction with the caregiver this robot was placed in a small cardboard box with the resources directly in front of it, in an attempt to create a sensory deprivation experience (see **Figure 5**). As would be expected, with both resources directly in front of it and little room to move, the robot remained mostly inactive throughout the sensory deprivation period as shown in **Figure 8**.

4. COMPARING THE COGNITIVE DEVELOPMENT OF OUR ROBOTS

From the overview of the experiments, it appears that the robot that developed in the novel environment (Section 3.2) gained more advanced cognitive abilities than the robots developed in the standard and "sensory deprivation" environments. These advanced cognitive abilities would seem to support the idea that an environment which provides a richer sensorimotor experience over the course of development leads to a greater cognitive development in autonomous robots too. However, we must ask the question whether these more advanced cognitive abilities are a permanent result of the actual developmental process, or a transient phenomenon due to the different environmental conditions. In order to try to understand if these developmental conditions had indeed affected the cognitive development of the robots, in the following section we compare the robots' neural networks and behavior in different developmental tests.

4.1. Comparison of Neural Development and Activity

For a first comparison between the robots, we will look in closer detail at the development of their different neural networks, which can be seen in **Figure 10**.

Figure 10 shows that the robot from the novel environment developed a larger neural network with significant growth occurring in the latter stages of the experiment, coinciding with the robot going through what in Section 3.2.2 we considered related to the coordination of secondary and tertiary reactions during the exploration period. Additionally, we can see again that the robot from the novel environment had a significantly larger number of neurons firing per action loop in the later stages. The increased number of nodes and neural activity from the novel robot can be explained due to this robot developing larger neural pathways. The increased pathways benefited the robot by giving it a better "understanding" of its environment.

4.2. Learning and Association: Introduction of a New Object

We next tested the ability of each of the three versions of the Koala robot – the robots from the experiments carried out in the "standard," "novel," and "sensory deprivation" environments – to learn by introducing two new novel objects that the robots had not seen before – two AIBO robots (one white and one black) shown in **Figure 5**. These novel objects were set to work in a similar manner to the energy resource, recharging the energy of the robot when it was close, although these novel objects provided a much greater rate of energy replenishment – 30 units of energy per second, 4 times faster than the original energy resource. The Koala robots were then given a choice between the novel objects and the original energy resource, with the assumption that if/once the robots learned that the novel resources provided a greater charge, they would prefer them over the original resources.

In order to conduct this experiment, two minor changes were made to the robot's architecture. First, the energy level was set to 20% after every action loop, to ensure the robot had a permanent motivation to recover from energy deficits. Second, the secretion of the nH_c was suppressed to remove the motivation to move to the novel resources based purely on their novelty value. The experiment involved two parts, with results show in **Table 2** and **Figure 11**.

For the first part of the experiment, the first novel object (the white AIBO) was placed directly in front of the Koala robot (close enough to charge) for a period of 10 s, to give the Koala an opportunity to learn about it; after this period, both this novel object and the original energy resource were placed slightly spread in front of the robot at a distance of around 1 m, forcing the robot to choose which one to move to in order to replenish its energy levels. This entire cycle was then repeated 10 times.

The results are reported in **Table 2**, where we can see that the "novel" robot appears to immediately learn the increased energy affordance provided by the first novel object and was significantly more attracted to it. In comparison, the "standard" robot would often pick the novel resource after increased exposure to it, although as seen in **Figure 11** it was only slightly preferred. The "sensory deprived" robot did not show any signs of adaptation, systematically selecting the original energy resource.

For the second part of this experiment, conducted immediately after the first part, we changed the first novel object with the second (the black AIBO). Unlike in the previous part of the experiment, the new novel object was not placed in front of the robot at any time; instead, it was placed 1 m ahead of the robot and slightly spread. Once again each of the versions of the Koala robot underwent another 10 runs with a similar need to replenish its energy level. While the robot had never seen the second novel object before, this object shares similarities with the first, hence here we are testing if the robot can identify that the two novel objects share similarities and therefore may behave in a similar manner, i.e., both would offer rapid replenishment of the energy deficit. The results are shown in **Table 2**, where we can see that even though the novel robot had never seen or interacted with the new novel object, unlike the other robots, due to its more
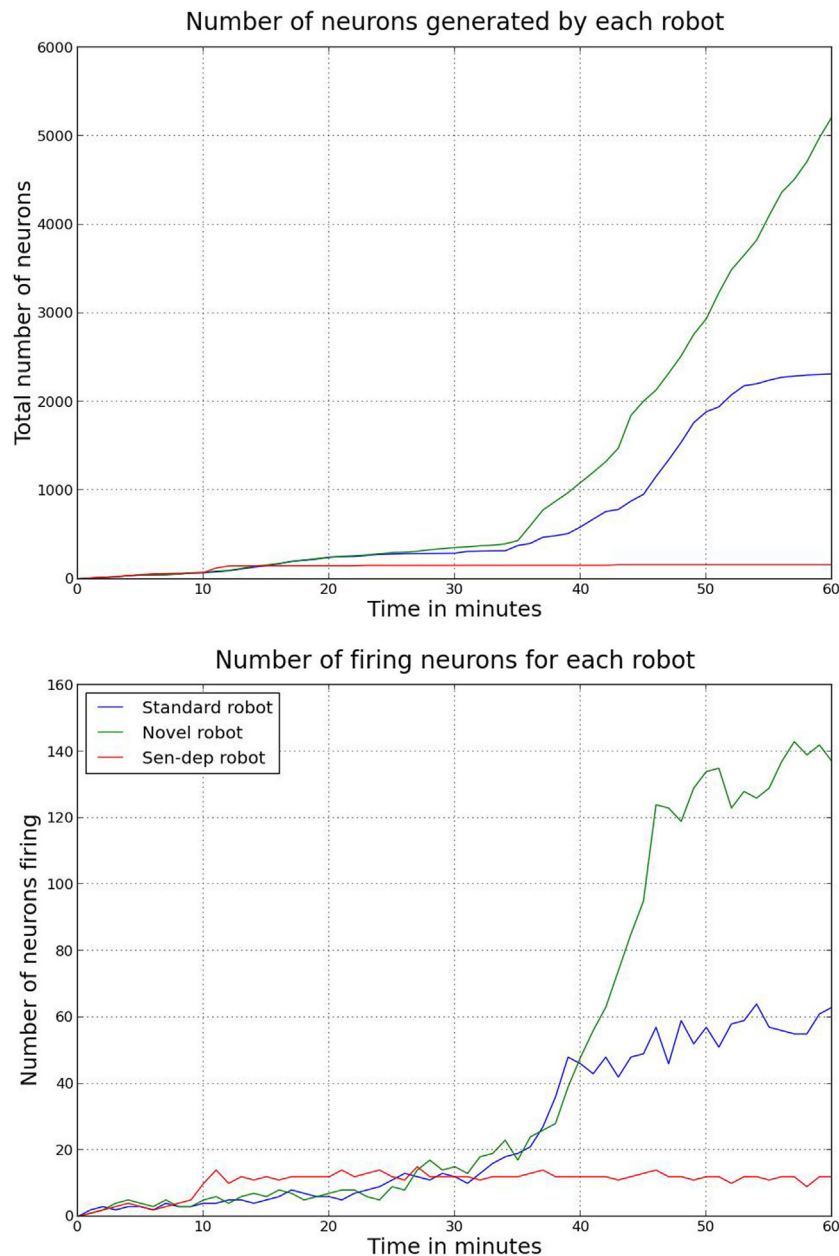
FIGURE 10 | The total number of neurons generated by the ENN for each of the robots over the course of the experiment (top) as well as the total number of firing nodes (bottom). The nodes themselves are generated as a function of the robots interaction with its environment. A higher number of nodes would suggest that the robot has learned about a larger number of, or in more detail about, different objects or aspects of its environment. A higher number of nodes firing appear to be related to the robot either noticing more aspects of the environment or having a greater understanding of the affordance of different aspects of the environment.

developed neural network (Section 4.1), it was able to identify the similarities between the two novel objects and recognize that the second had similar properties (i.e., the ability to provide a rapid charge) to the first. In contrast, while the standard robot did seem to identify some similarities between the two novel objects, leading to a slight perceived affordance of energy recovery with the new objects, the perception was not good enough for it to choose the novel object over the safer original energy resource.

Finally, the sensory deprived robot, which only showed minimal learning in the first stage of this experiment, showed no association between the two objects.

4.3. Object Permanence: Recreation of a Hidden-Toy Test

One of the tests most commonly used in developmental psychology to assess whether infants have acquired the notion

TABLE 2 | The robots' choices between the novel and the original resource in the first (left) and the second part (right) of the learning experiment.

	First experiment				Second experiment		
Run	Standard robot	Novel robot	Sensory-dep robot	Run	Standard robot	Novel robot	Sensory-dep robot
1	Original	Novel	Original	1	Original	Novel	Original
2	Original	Novel	Original	2	Original	Novel	Original
3	Novel	Novel	Original	3	Original	Novel	Original
4	Novel	Novel	Original	4	Original	Novel	Original
5	Novel	Novel	Original	5	Original	Novel	Original
6	Novel	Novel	Original	6	Original	Novel	Original
7	Novel	Novel	Original	7	Original	Novel	Original
8	Original	Novel	Original	8	Original	Novel	Original
9	Novel	Novel	Original	9	Original	Novel	Original
10	Novel	Novel	Original	10	Original	Novel	Original

of object permanence is the hidden toy test[4] (Piaget, 1952; Baillargeon, 1993; Munakata, 2000). We reproduced this test by placing a needed resource in front of each of the Koala robots at a range of 2 m. As the robot began to move toward the resource, 5 tins cans, used to build the previous novel structure shown in **Figure 5**, were placed directly in front of the resource to block it from the robot's view. If the robot has a representation of object permanence, we would hypothesize that the robot would continue to move toward the object even when it is hidden from sight. If the robot stopped for more than 10 seconds, or 1 minute had passed after the resource had been hidden, the experiment ended to reduce the risk that the robot might find the resource accidentally or as part of its exploratory behavior. This experiment was conducted 10 times for each robot, and the results are shown in **Table 3**.

As shown in **Table 3**, the robots from the novel and standard environments both had some success in finding the resources once hidden; in comparison, the robot from the sensory deprivation environment was unsuccessful every time. If we look at the behavior and neural activity of the robots, shown in **Figure 12**, we can see that the robot from the novel environment was the only one to consistently search for the resource after it was hidden. In addition, this robot was also the only one which consistently (i.e., in every run) had high activity along the neural pathway associated with the detection of the resource even after it had disappeared. This neural activity resulted from the fact that the original signal remained active along the pathway due to the modulation of this pathway by the different hormone concentrations, leading to feedback loops. These loops provide the robot with an ability akin to "active," or "short term" memory.

The 3 occasions when this robot failed to find the resource were due to the fact that the robot moved past the hidden resource. The fact that the neural activity remained high during these failed attempts suggests that, while the robot shows neural

---

[4]We have also carried out experiments using the A-not-B test with robots that were developed in an environment with human "caregivers". Results of these experiments will be reported in a forthcoming publication.

activity associated with the hidden object and the behavior it affords, without the expected feedback from sensory readings regarding the distance and position of the object, the robot cannot consistently locate it. This would appear to back up the previous observation that the first two robots had gained an ability consistent with the "understanding" of object permanence during their developmental runs, rather than having this skill from the start.

4.4. Violation of Expectation Paradigm

For the final experiment, we tested the robots using another common cognitive test, the Violation of Expectation paradigm (VOE). VOE experiments are normally carried out by showing very young infants two different pictures, one of which shows an impossible outcome – often some type optical illusion – while the other is almost identical but without the impossibility (Sirois and Mareschal, 2002). The experiment seeks to assess if the baby can notice the impossibility by measuring which picture it looks at more. The underlying assumption is that if the baby can identify the impossibility in the picture, it must have some expectation about the object represented in that picture, and will look at it for longer than at the image without the impossibility.

We created a version of this experiment suitable for our robots. Here, a white ball was placed in front of the robot. For the possible outcome, we simply measured how long the ball which has not been seen before held the robot's attention. For the "VOE," the white ball was again placed in front of the robot; however, the robot's sensors were manipulated to make it appear as if the ball became smaller as the robot moved toward it. We once again measured how long the ball held the robot's attention. If the robot can identify the "VOE," we would expect it to hold its attention for a longer period of time.

As shown in **Table 4**, the "VOE" held the novel robot's attention for significantly longer than the possible object. In contrast, the robot from the standard environment only showed slightly more interest in the "VOE"; due to the small difference, it is not possible to conclusively suggest that this robot was showing an interest in the "VOE." Finally, the sensory deprived robot showed no real difference in the time spent with both objects, suggesting the VOE paradigm had no real influence on the robot.

Our results suggest that the ability to respond to "VOE" arises as part of the later stages of the sensorimotor development process, which only the robot from the novel environment went through. Incidentally, this finding correlates with Piaget's developmental theory regarding when these skills should emerge. It should be noted that the VOE paradigm has been criticized by other developmental psychologists, who debate whether these skills are indeed learned as suggested by Piaget (1952) in his theory of development, or if they are part of a core knowledge that all infants possess from birth, as suggested by Baillargeon et al. (1985), Baillargeon (1993), and Spelke et al. (1992). These authors have previously used the VOE paradigm to demonstrate that babies are able to identify impossibility much earlier than would be expected if Piaget's theory was correct. In the case of a robot, we can be certain that this was not part of the robot's core knowledge but is developed given the appropriate sensorimotor experiences.
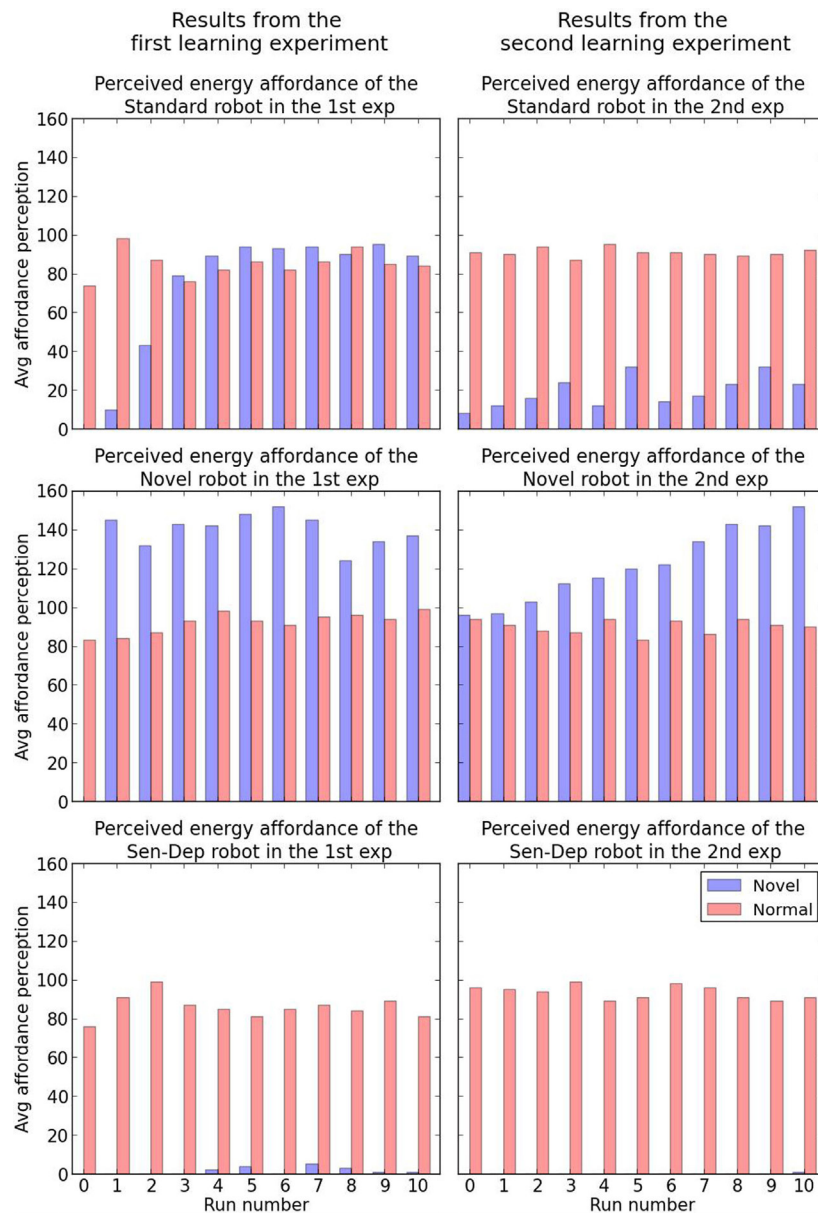
FIGURE 11 | **The results of the first (left) and second (right) part of the learning experiment**. The units in the y-axis show the strength of the perceived energy affordance of the two objects which is determined by equation (12). Run 0 is the perceived energy affordance before the start of experiments.

TABLE 3 | **Results of the hidden-toy test.**

Run	Novel robot's time	Standard robot's time	Sens-dep robot's time
1	32	16	Not found
2	14	Not found	Stopped
3	17	Not found	Stopped
4	32	32	Stopped
5	Not found	Not found	Stopped
6	Not found	Not found	Stopped
7	15	Not found	Stopped
8	Not found	43	Not found
9	23	Not found	Stopped
10	19	12	Stopped

Time is measured in seconds.

There have also been debates (Baillargeon et al., 1985; Baillargeon, 1993) regarding whether the violation of expectation paradigm used with infants may have been biased, suggesting that the impossible variation offered other additional stimuli (e.g., more activity or increased number of elements in the impossible picture), which attracts the infants' attention rather than their ability to identify or be attracted to the perceived impossibility. We could also imagine that the higher novelty of the impossible object might to some extent be responsible for the response of the infants. In the case of our robot experiments, we would be happy to accept that the different response that the novel robot displays in the face of impossible experiences might be due to novelty. However, the novelty offered by the impossible

FIGURE 12 | Overview of the behavior and neural activity of our three robots during the hidden toy test. The graph on the left shows the neural activity along the pathways associated with the hidden resource of each robot. Neural activity is measured as the percentage of active nodes. Crosses indicate the points at which each robot has found (i.e., detected) the hidden resource. Dotted lines are used to show the neural activity of the robot after this point. In most cases, the robot will detect a resource some seconds before it physically reaches it. As we can see, the perception of the hidden resource gives rise to increased neural activity. After the resource is hidden (at around the 10s mark), the pathway of the robot from the novel environment is the most active even without being able to see the resource: this robot has neural activity associated with the object and the behavior that it affords even after it has been hidden from view. On the right, the diagrams show the trajectories of the robots for each run. We can see that the robot from the novel environment was the most successful in finding the hidden resource.

TABLE 4 | Results from the VOE experiment showing the time (in seconds) spent focusing on or interacting with the impossible and possible object.

Run	Novel robot's time		Base robot's time		Sens-dep robot's time	
	Impossible	Possible	Impossible	Possible	Impossible	Possible
1	46	17	21	18	19	20
2	48	15	22	19	23	21
3	43	17	19	22	21	23
4	49	18	17	15	20	20
5	54	14	19	12	15	15
6	52	18	16	15	19	20
7	49	19	15	14	18	19
8	56	16	14	13	19	20
9	49	14	16	16	14	13
10	54	17	21	15	18	16

object (a white ball which behaves in a way that violates all the robot's previous sensorimotor experiences) is very different from the type of novelty offered by the perception of the novel object (white ball).

Although both exposure to a new object and a "VOE" episode produce "novelty," which we could define as the lack of behaviors and representations associated with an object, there is a qualitative difference in the effects that both experiences have on the robot's neural network. Novelty that arises from exposure to a new object is related to the level of plasticity between two connecting neurons [see equation (15)]: the higher the novelty of the object, the lower the level of plasticity, and a totally new object will give rise to new nodes and connections. Novelty related to

a violation of expectation episode involves, in addition to the above, a change in the neural pathway associated with closely related previous experience. Specifically, an existing pathway is activated but a number of new nodes rapidly emerge along this pathway, linked to the elements of the new experience that violate the expectations from previous experiences (see **Figure 13**, label "Novel Robot"). Due to the activity of the pathway and the rapid growth of new nodes along it, a large number of "messy" and overlapping connections between nodes are quickly generated, increasing perceived novelty due to their high plasticity. These overlapping connections also effectively lead to the emergence of a sort of positive feedback loop within the pathway. This results in a greater level of activity along that pathway and therefore increased novelty. Over time, if the robot is repeatedly exposed to the "violation of expectation," the original nodes and the new nodes will separate along two distinct pathways, the overlapping connections get "pruned," and the feedback loops disappear (see Section 2.3.2), as shown in **Figure 13**, label "Novel Robot after repeated exposure". This results in the consolidation of the new pathway and thus the previous "violation of expectation" becomes a normal "experience" for the robot. However, if exposure to the "VOE" stimulus is infrequent, the pathways will not split and the novelty associated with it will persist.

5. CONCLUSION AND DISCUSSION

In this paper, we have demonstrated the importance of sensorimotor experiences and environmental conditions in the emergence

FIGURE 13 | A snapshot of the neural pathway associated with a "violation of expectation episode. The greater effect can be seen in the network of the "Novel Robot": exposure leads to the generation of a larger number of nodes and connections leading to a greater neural activity and to the creation of a new pathway if exposure to the same type of episode re-occurs often enough.
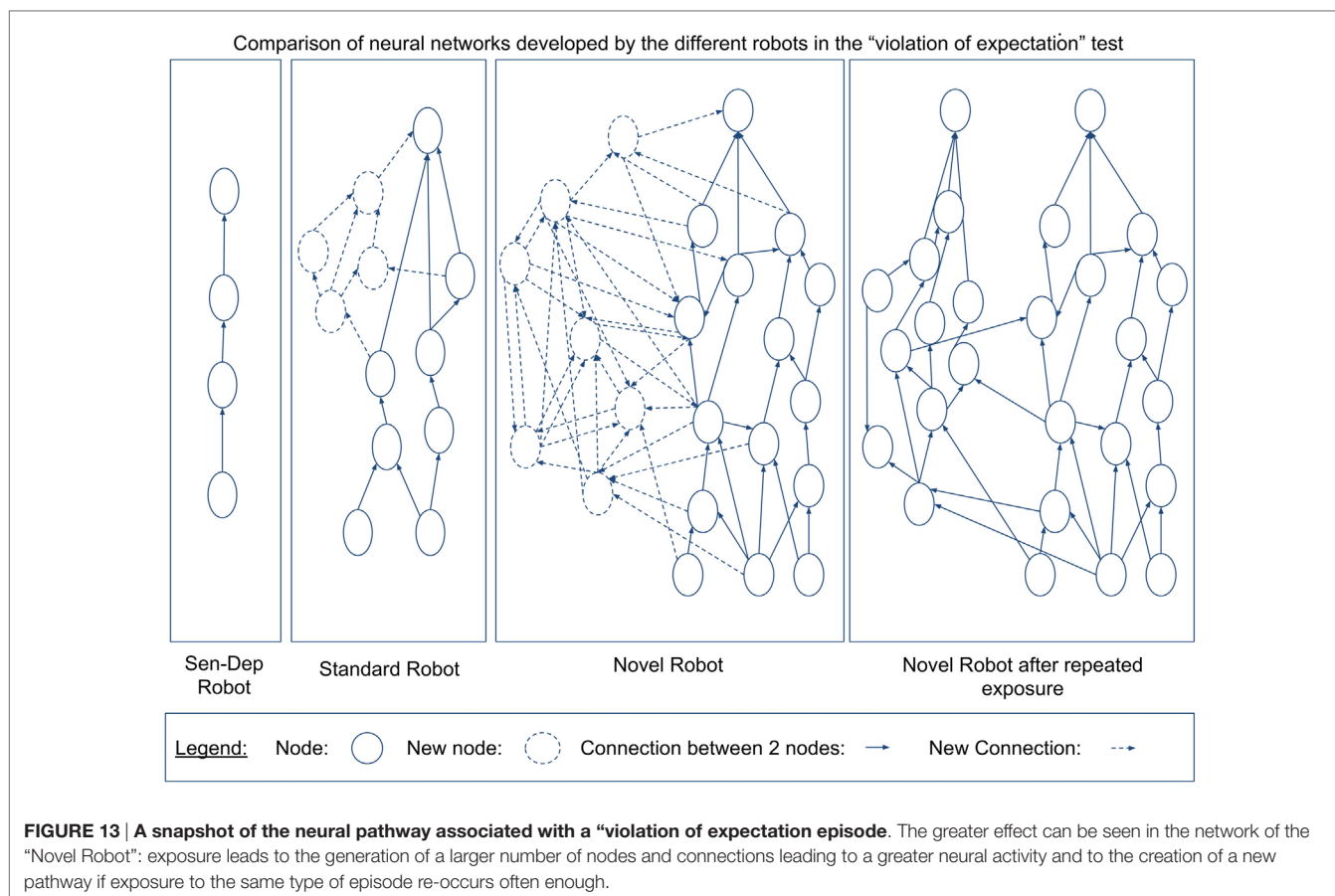
of more advanced cognitive abilities in an autonomous robot. The robot exposed to a wide range of novel sensorimotor experiences and different stimuli showed greater cognitive abilities than the robots from the other environments. Particularly, the robot raised in a context of sensory deprivation showed no additional behaviors or abilities outside simple reflex-like behavior that each robot started with. Furthermore, this sensory deprived robot performed badly at adapting and in learning tasks compared with the other robots. Our model thus shows that a richer sensorimotor experience during early development correlates with greater cognitive ability.

We have also shown how an autonomous robot implementing an epigenetic architecture has the potential to go through developmental stages in a similar manner as outlined in Piaget's sensorimotor theory. Our robot starts with a simple reflex-like behavior, yet through interactions with the external environment and stimulation from its internal environment, it develops more complex behaviors and cognitive abilities. Our robot was not explicitly designed around a developmental theory, but these developmental substages emerged purely due to the interactions among the different aspects of the architecture – the hormonal, epigenetic, and ENN – and the external environment. In our past studies such as Lones and Cañamero (2013), when not all the previous components were present, the developmental phenomena described here did not emerge. In particular, the addition of the ENN, with its learning and

representational capabilities, in interaction with the other elements of the architecture, was a key factor in the emergence of these developmental phenomena. Our model thus offers potentially useful insights to bridge gaps between studies of epigenetic mechanisms such as Crews' (2010) and developmental epigenetic theories such as Piaget's (1952) by showing how the former can lead to the emergence of the latter.

AUTHOR CONTRIBUTIONS

JL: lead author, designed, conducted, and analyzed the experiments, wrote the first draft of the paper, and addressed the revisions. ML: second Ph.D. supervisor of the lead author, provided substantial advice regarding formalization and analysis of the experiments and substantial contributions to improving, editing, and rewriting various drafts and proofs of the paper. LC: main Ph.D. supervisor of the lead author and senior author of the paper, provided general advice on this work and substantial contributions to improving, editing, and rewriting various drafts and proofs of the paper.

FUNDING

REFERENCES

Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., et al. (2009). Cognitive developmental robotics: a survey. *IEEE Trans. Auton. Ment. Dev.* 1, 12–34. doi:10.1109/TAMD.2009.2021702

Avila-Garcia, O., and Cañamero, L. (2005). "Hormonal modulation of perception in motivation-based action selection architectures," in *Procs of the Symposium on Agents that Want and Like* (Hatfield: SSAISB), 9–17.

Bahrick, L. E., Lickliter, R., and Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Curr. Dir. Psychol. Sci.* 13, 99–102. doi:10.1111/j.0963-7214.2004.00283.x

Baillargeon, R. (1993). The object concept revisited: new directions in the investigation of infants physical knowledge. *Vis. Percept. Cogn. Infancy* 23, 265–315.

Baillargeon, R., Spelke, E. S., and Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition* 20, 191–208. doi:10.1016/0010-0277(85)90008-3

Baird, A. A., Kagan, J., Gaudette, T., Walz, K. A., Hershlag, N., and Boas, D. A. (2002). Frontal lobe activation during object permanence: data from near-infrared spectroscopy. *Neuroimage* 16, 1120–1126. doi:10.1006/nimg.2002.1170

Berridge, K. C. (2004). Motivation concepts in behavioral neuroscience. *Physiol. Behav.* 81, 179–209. doi:10.1016/j.physbeh.2004.02.004

Blanchard, A., and Cañamero, L. (2006). "Developing affect-modulated behaviors: stability, exploration, exploitation or imitation," in *Proc. 6th Intl. Workshop on Epigenetic Robotics*. Lund University Cognitive Studies (LUCS) Vol. 128, 17–24.

Breazeal, C. (1998). "A motivational system for regulating human-robot interaction," in *The Fifteenth National Conference On Artificial Intelligence* (Madison: Aaai/iaai), 54–61.

Cañamero, D. (1997). "Modeling motivations and emotions as a basis for intelligent behavior," in *Proceedings of the First International Conference on Autonomous Agents* (Marina del Rey, CA: ACM), 148–155.

Cañamero, L., Avila-Garcia, O., and Hafner, E. (2002). "First experiments relating behavior selection architectures to environmental complexity," in *Procs IEEE/RSJ Int Conf on Intelligent Robots and Systems* (Lausanne: IEEE).

Cangelosi, A., and Riga, T. (2006). An embodied model for sensorimotor grounding and grounding transfer: experiments with epigenetic robots. *Cogn. Sci.* 30, 673–689. doi:10.1207/s15516709cog0000_72

Cangelosi, A., Schlesinger, M., and Smith, L. B. (2015). *Developmental Robotics: From Babies to Robots*. Cambridge, MA: MIT Press.

Cos, I., Cañamero, L., and Hayes, G. M. (2010). Learning affordances of consummatory behaviors: motivation-driven adaptive perception. *Adapt. Behav.* 18, 285–314. doi:10.1177/1059712310375471

Crews, D. (2008). Epigenetics and its implications for behavioral neuroendocrinology. *Front. Neuroendocrinol.* 29:344–357. doi:10.1016/j.yfrne.2008.01.003

Crews, D. (2010). Epigenetics, brain, behavior, and the environment. *Hormones (Athens)* 9, 41–50. doi:10.14310/horm.2002.1251

DeLoache, J. S. (2000). Dual representation and young children's use of scale models. *Child Dev.* 71, 329–338. doi:10.1111/1467-8624.00148

Di Paolo, E. A. (2003). "Organismically-inspired robotics: homeostatic adaptation and teleology beyond the closed sensorimotor loop," in *Dynamical Systems Approach to Embodiment and Sociality, Advanced Knowledge International* (Adelaide: Citeseer), 19–42.

Fischer, K. W. (1980). A theory of cognitive development: the control and construction of hierarchies of skills. *Psychol. Rev.* 87, 477. doi:10.1037/0033-295X.87.6.477

Fuster, J. M. (2002). Frontal lobe and cognitive development. *J. Neurocytol.* 31, 373–385. doi:10.1023/A:1024190429920

Holliday, R. (2006). Epigenetics: a historical overview. *Epigenetics* 1, 76–80. doi:10.4161/epi.1.2.2762

Kagan, J., and Herschkowitz, N. (2006). *A Young Mind in a Growing Brain*. New York: Psychology Press.

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT press.

Krichmar, J. L. (2008). The neuromodulatory system: a framework for survival and adaptive behavior in a challenging world. *Adapt. Behav.* 16, 385–399. doi:10.1177/1059712308095775

Krichmar, J. L. (2013). A neurorobotic platform to test the influence of neuromodulatory signaling on anxious and curious behavior. *Front. Neurorobot.* 7:1. doi:10.3389/fnbot.2013.00001

Lones, J., and Cañamero, L. (2013). "Epigenetic Adaptation through Hormone Modulation in Autonomous Robots," in *2013 IEEE 3rd Joint Int Conf on Development and Learning and Epigenetic Robotics*. Osaka: IEEE.

Lones, J., Lewis, M., and Cañamero, L. (2014). "Hormonal modulation of development and behaviour permits a robot to adapt to novel interactions," in *ALIFE 14: The Fourteenth Conference on the Synthesis and Simulation of Living Systems*, Vol. 14. New York, 184–191.

Marshall, J., Blank, D., and Meeden, L. (2004). "An emergent framework for self-motivation in developmental robotics," in *Proceedings of the 3rd International Conference on Development and Learning (ICDL 2004)* (San Diego: Salk Institute), 10.

Munakata, Y. (2000). Challenges to the violation-of-expectation paradigm: throwing the conceptual baby out with the perceptual processing bathwater? *Infancy* 1, 471–477. doi:10.1207/S15327078IN0104_7

Oudeyer, P.-Y., and Smith, L. (2014). How evolution may work through curiosity-driven developmental process. *Topics Cogn. Sci.* 8, 492–502.

Papalia, D. E., Olds, S. W., and Feldman, R. D. (1992). *Human Development*. New York: McGraw-Hill.

Piaget, J. (1952). *The Origins of Intelligence in Children*. New York: International University Press.

Piaget, J., and Inhelder, B. (1969). *The Psychology of the Child*. New York: Basic Books.

Schaffer, H. R. (1996). *Social Development*. Oxford: Blackwell Publishing.

Schlichting, C. D., and Pigliucci, M. (1998). *Phenotypic Evolution: A Reaction Norm Perspective*. Sunderland, MA: Sinauer Associates Incorporated.

Shaw, P., Law, J., and Lee, M. (2012). An evaluation of environmental constraints for biologically constrained development of gaze control on an icub robot. *Paladyn* 3, 147–155. doi:10.2478/s13230-013-0103-y

Sirois, S., and Mareschal, D. (2002). Models of habituation in infancy. *Trends Cogn. Sci.* 6, 293–298. doi:10.1016/S1364-6613(02)01926-5

Spelke, E. S., Breinlinger, K., Macomber, J., and Jacobson, K. (1992). Origins of knowledge. *Psychol. Rev.* 99, 605. doi:10.1037/0033-295X.99.4.605

Timmis, J., Murray, L., and Neal, M. (2010). "A neural-endocrine architecture for foraging in swarm robotic systems," in *NICSO 2010: Nature Inspired Cooperative Strategies for Optimization*, Vol. 284. eds J. R. Gonzalez, D. A. Pelta, C. Cruz, G. Terrazas, and N. Krasnogor (Springer-Verlag Berlin), 319–330.

Ugur, E., Nagai, Y., Sahin, E., and Oztop, E. (2015). Staged development of robot skills: behavior formation, affordance learning and imitation with motionese. *IEEE Trans Auton Ment Dev.* 7, 119–139.

Zhang, X., and Ho, S.-M. (2011). Epigenetics meets endocrinology. *J. Mol. Endocrinol.* 46, R11–R32. doi:10.1677/JME-10-0053

Behavioral Diversity Generation in Autonomous Exploration through Reuse of Past Experience

Fabien C. Y. Benureau[1,2,3]* and Pierre-Yves Oudeyer[1,2]

[1] FLOWing Epigenetic Robots and Systems (FLOWERS) Team, Inria Bordeaux Sud-Ouest, Bordeaux, France, [2] École Nationale Supérieure de Techniques Avancées (ENSTA) ParisTech, Palaiseau, France [3] Bordeaux University, Bordeaux, France

The production of behavioral diversity – producing a diversity of effects – is an essential strategy for robots exploring the world when facing situations where interaction possibilities are unknown or non-obvious. It allows to discover new aspects of the environment that cannot be inferred or deduced from available knowledge. However, creating behavioral diversity in situations where it is most crucial – new and unknown ones – is far from trivial. In particular in large and redundant sensorimotor spaces, only small areas are interesting to explore for any practical purpose. When the environment does not provide clues or gradient toward those areas, trying to discover those areas relies on chance. To address this problem, we introduce a method to create behavioral diversity in a new sensorimotor task by re-enacting actions that allowed to produce behavioral diversity in a previous task, along with a measure that quantifies this diversity. We show that our method can learn how to interact with an object by reusing experience from another, that it adapts to instances of morphological changes and of dissimilarity between tasks, and how scaffolding behaviors can emerge by simply switching the attention of the robot to different parts of the environment. Finally, we show that the method can robustly use simulated experiences and crude cognitive models to generate behavioral diversity in real robots.

Keywords: exploration, transfer learning, sensorimotor, robot, behavioral diversity

1. MOTIVATION

The engagement of robots and animals with the world generates a complex sensorimotor flow, which features a large motor space and multiple sensory modalities. While the body, as an active interface to the environment, simplifies in important ways the raw experience of the world (Hoffmann and Pfeifer, 2012), the learning and decision-making challenges the individual faces are still formidable.

In recent years, the *child-as-a-scientist* paradigm (Gopnik, 1997, 2012; Schulz and Bonawitz, 2007; Gweon and Schulz, 2008; Cook et al., 2011) has emerged as a major paradigm of child cognitive development. It considers the hypothesis that children can act as would rational thinkers, creating experiments and testing hypotheses through their interaction with the world in a manner structurally similar to scientific inquiry. Several works have indeed shown that preschoolers understand causality, distinguish it from spurious associations, and construct interventions to do so (Gopnik et al., 2001; Schulz et al., 2007).

Constructing and carrying out informative interventions, i.e., interactions that afford information gain, decrease the number of interactions necessary to understand a phenomenon, therefore

ensuring an economy of time and energy. Yet, it also requires cognitive resources that may either be lacking (the individual cannot grasp the situation with his current cognitive abilities) or may represent too high an effort to justify the information gain they afford.

Robots – the focus of this paper – face a similar situation. In autonomous developmental contexts, robots do not have access to descriptions of their environments crafted by experts. Rather, they have to learn from experience. When social peers are not available, this experience has to be acquired autonomously from their own exploration of the environment. They have to act in unfamiliar situations that – due to a fundamental scarcity of knowledge – escape, at first, their abilities to fully grasp them, either through representation, prediction, control, or planning.

Designing informative interventions in those situations then faces a chicken-and-egg problem: knowing which interventions are going to be informative requires information that is not yet available, and that must be acquired through informative interventions.

Of course, that does not mean that informative interventions cannot be conducted, as any interaction can turn out to be informative *a posteriori*. But the fundamental problem of choosing which interventions to conduct while being unable to predict which ones are going to be informative remains.

A possible strategy, then, is to create *behavioral diversity*. Behavioral diversity characterizes the number and variety of behaviors an agent exhibits in its environment. Determining how different two behaviors are largely depends on the observer and its motives. For instance, a humanoid robot placed on a surface and executing random motor activations will engage in complex and unique patterns of movements and will end up convulsing on the floor most of the time. Arguably, each pattern of movement can be considered as a different behavior. But for a task such as standing up, the elevation of the head during movement might be the only relevant signal. In that perspective, all patterns of movement resulting in convulsions on the floor represent the same behavior while the robot sitting up or standing up represent different ones.

In this paper, and in the context of an autonomous robotic perspective, we characterize behaviors through the *environmental feedback* they elicit, as perceived by the robot itself, rather than the actions they necessitate. Exhibiting behavioral diversity then equates producing a diversity of effects in the environment. The dimensions of those define what we will call here, with three interchangeable terms: behavioral space, effect space, or sensory space.

Producing behavioral diversity can be a good strategy in unknown situations because it is not directed toward – and as such not constrained to produce information about – a specific goal. Instead, it creates a set of observations about diverse features of the environment, offering the robot a set of options that can be explored and exploited toward specific objectives afterward. The usefulness of such strategies for robots has been recently explored through models of curiosity-driven learning and intrinsically motivated reinforcement learning (Oudeyer and Kaplan, 2007; Baldassarre and Mirolli, 2013; Benureau and Oudeyer, 2015), and in a related line of work, on novelty and diversity search in evolutionary robotics (Mouret and Doncieux, 2009; Lehman and Stanley, 2011).

In many practical contexts, the situation facing the robot is not completely unknown, and rational deductions and inferences can be made about what type of interactions are going to be informative. Still, they may not narrow the number of candidate interventions to a reasonable number. In that context, producing behavioral diversity can be seen as an essential strategy to deal with the limits of pure logical reasoning. It provides a heuristic mechanism for discovering knowledge by the learner-as-a-scientist (be it a human or a robot) when rational mechanisms used to uncover the laws of the world cannot be applied.

In other words, such a heuristic picks up when rational deductions end: logical reasoning identifies a set of interactions worth trying, and the behavioral diversity heuristic provides a sampling method to choose what to try among this remaining set. For instance, a child playing with a teddy bear and a rattle may understand that to figure out what a rattle does, interactions with the teddy bear are uninformative. This halves the space of candidate interactions but provides no clue about which interactions are interesting to try on the rattle. Trying to interact with the rattle in different ways is then an effective strategy.

This selective exploration principle is elegantly formulated by Cook et al. (2011) in the children's case: "selective exploration of confounded evidence is advantageous even if children explore randomly (with no understanding of how to isolate variables): the more different actions children perform, the better their odds of generating informative data." (p. 352). Gweon and Schulz (2008) provide a study where children presented with confounding evidence increase the variability of their exploration, even if that represents a physical effort. Schulz and Bonawitz (2007) and Bonawitz et al. (2012) report similar results: children preferentially engage with a confounding toy, rather than to play with a new one.

But producing behavioral diversity is not necessarily a trivial task. While producing random motor behavior is algorithmically straightforward – it boils down to picking a random motor action among the ones available – it does not typically generate a diversity of interactions and effects in the environment: motor diversity does not translate into effect diversity. This is caused by the typical heterogeneous distribution of the redundancy of the sensorimotor space: to some effects correspond a large number of motor commands, while some effects can only be produced by a small, specific set of them. In an interaction task with an object, for instance, few motor commands may actually produce contact with the object. The rest of them produce the same effect on the object: nothing. As such, a uniform, random sampling of a large motor space will only produce effects in highly redundant parts of the sensory space for any reasonable (i.e., at the timescale of a lifetime) number of samples. Therefore, an efficient sampling strategy must be devised.

However, as producing behavioral diversity is most useful in tasks where little or no knowledge of the underlying environmental mechanisms exists, and this knowledge is precisely what would be needed to choose which interactions to carry in order to produce effects as diverse as possible, the production of behavioral diversity suffers a similar chicken-and-egg problem as the one raised by the design of informative interventions: it needs knowledge to create interactions that will generate data that will serve to derive the knowledge it needs.

One possibility to break the circularity is to procure knowledge from somewhere else. In this paper, we introduce a method to create behavioral diversity in an unknown task by leveraging past experience from another task. We consider a scenario where one task has been explored, and a new, unknown task is presented to the robot. The relationship between the two tasks is not given to the robot, and can be arbitrary. The only constraint is that at least some motor commands executed in the previous task can be reexecuted in the new one. Besides this, the method transparently adapts to arbitrary changes in sensory modalities and learning algorithms between tasks.

In the next section, we first formalize the problem (Section 2) and introduce a measure to quantify behavioral diversity in continuous sensorimotor spaces. We then introduce our method (Section 3). In Section 4, we present a simple application of the method. Then, in Section 5, we detail a more complex situation where a robotic arm interacts with different objects.

2. PROBLEM

An *environment* is here formally defined as a mapping f from M to S, which can be stochastic. M is the motor space, and it represents a parameterization of the movements the robot can execute. It is a bounded hyperrectangle of $\mathbb{R}^{d_M}$; d_M is the dimension of the motor space. S is the effect space, of dimension d_S; it is a bounded subset of $\mathbb{R}^{d_S}$. *Effects* and *goals* (i.e., desired effects) are elements of S. In this paper, both M and S are multidimensional continuous spaces, with $d_S \ll d_M$.

Here, the elements of the motor space do not directly encode the raw commands that the motors of the robot receive. Instead, we use *motor primitives* that transform vectors of parameters from the motor space M into streams of real-time, hardware-specific motor commands. A motor primitive can be a simple goal position for a given motor, or be a Dynamic Movement Primitive (DMP) (Ijspeert et al., 2013) that translates parameters into smooth motor trajectories; both will be used in this paper. Likewise, the sensory space does not contain the raw readings of the sensors but rather *behavioral descriptors*: parameterized behavioral representations of raw sensors data after it has been processed by *sensory primitives*. Concretely, a sensory primitive can encode the position of an end-effector in Cartesian space or the displacement of an object after a robot interacted with it. This allows to flexibly encode sensory feedback into high-level representations. Such sensory primitives do not only abstract low-level feedback data: they represent the robot's attention, by encoding specific features of the environment and not others, and we use them deliberately this way in this paper.

Environments are black boxes, and only the parameterizations M and S are known to the robot. Let us remark that, while valuable information can be encoded in the boundaries of S, nothing prevents S to be arbitrarily large compared to the *reachable space* $f(M)$, i.e., the set of effects that can actually be produced. In order to avoid unnecessary complexity in this paper, we will only consider experiments where S is not significantly larger than the axis-aligned bounding box of $f(M)$. A method to deal with arbitrarily large S can be found in Benureau and Oudeyer (2015).

An *exploration task*, subsequently referred simply as a *task*, is defined as a pair (f, n) with $f: M \mapsto S$ the environment and n the maximum number of samples of f allowed, i.e., the number of actions the robot can execute in the environment.

We will consider scenarios made of two tasks, a task $A = (f_A, n_A)$, the *source* task, and a task $B = (f_B, n_B)$, the *target* task. We assume that motor commands from M_A can be reexecuted in the target task. In this paper, we will consider $M_A = M_B$, but other scenarios are possible, such as the existence of a known mapping between the two motor spaces (for instance, reusing motor commands used on the left arm of a humanoid on its right one). The reexecutability constraint is a strong one, but as robots body typically change much less quickly than their environments, many tasks share the same motor space. This may be less true for high-level motor, or action, spaces, but if no known mapping exists for the action spaces of two different tasks, the method does not just faces a problem of applicability: it is also probably of little use.

The source task is considered to have been interacted with using an arbitrary method, generating a sequence of observations $\{\mathbf{x}_i, \mathbf{y}_i\}_{0 \leq i < n_A}$ in $(M_A \times S_A)^{n_A}$, composed of the executed motor commands and observed effects. On the other hand, the robot has not yet interacted with the target task B.

The problem we are tackling in this paper is the question of transfer: how can the previous interaction with task A can be exploited to improve the exploration of task B?

We compare the case where information from A is exploited versus the situation where it is not, using as a baseline mechanism a random goal babbling architecture (SAGG-Random). Goal babbling has previously been shown to be an efficient strategy for the acquisition of inverse models (Baranes and Oudeyer, 2013; Moulin-Frier and Oudeyer, 2013) and in the production of behavioral diversity. We compare both cases using a behavioral diversity measure: *threshold coverage* (Benureau and Oudeyer, 2015). Improving the exploration of task B therefore means increasing the threshold coverage.

2.1. Threshold Coverage

Threshold coverage or τ-coverage is a behavioral diversity measure: it considers only the consequences of the motor commands, i.e., in our autonomous context, the behavioral effects as encoded in S, not the motor activations themselves. Motor *motions* can of course be part of behavior and contribute to diversity, when adequate sensors and sensory primitives are used to encode them in S. This is, for instance, the case in the behavioral descriptors used by the MAP-Elites algorithms (Cully et al., 2015).

Threshold coverage considers the volume of the union of the set of hyperballs of radius τ – the threshold – that have for centers the observed effects (**Figure 1**).

Formally, considering a set of points C belonging to $\mathbb{R}^n$, and $\tau \in \mathbb{R}^+$, we define the τ-coverage of C as:

$$\text{coverage}_\tau(C) = \text{volume}\left(\bigcup_{\mathbf{y}_i \in C} B(\mathbf{y}_i, \tau)\right)$$

with $B(\mathbf{y}_i, \tau)$ the hyperball of center $\mathbf{y}_i$ and radius τ.

The threshold coverage measure allows to quantify how much of the effect space is not more distant than τ of an observed effect.

FIGURE 1 | In this 2D example, the threshold diversity is represented by the green area (top). The threshold diversity measure quantifies how spread are the observed effects, up to a threshold τ. The measure is not sensitive to differences between distributions where all centers are pairwise more distant than τ (bottom).

FIGURE 2 | In this schematic representation, four clusters of effects are produced through different types of redundancies. From top to bottom, the first two effects are similar because their motor commands are similar. The second cluster exhibits body redundancy: two different motor commands end up generating the same forces on the environment. The third exhibits environmental redundancy: different forces produce the same effect. The fourth exhibits all of the three previous cases. Assuming the environment is neither stochastic nor chaotic, when selecting a diverse set of effects in the sensory space (for instance, the set highlighted in orange), the set of motor commands that generated them tends to display low body redundancy.

As a consequence, the threshold coverage measure is insensitive to differences between sets where the observed effects are pairwise more distant than τ (**Figure 1**).

Computing the threshold coverage requires to compute the volume of an arbitrary set of balls of the same radius. Exact methods exist using Voronoi Power Diagrams (Cazals et al., 2011), that partition the space into as many areas as there are balls; in each area, the center of only one ball is present, and the contribution of this ball to the overall volume can be computed independently of the others. There are also approximate methods based on Monte-Carlo sampling (Till and Ullmann, 2009).

We use the threshold coverage to characterize and contrast the robots' behavior under different algorithms. Let us remark that the robot, as an autonomous agent, never has access to the threshold coverage measure; it is purely an experimenter's tool.

3. METHOD

The idea behind our algorithm is to select a subset of motor commands executed during the exploration of the source task, and reexecute each of them on the target task. This subset is assembled with motor commands that generated diverse effects, i.e., that generated behavioral diversity, in the source task.

The assumption is that the production of behavioral diversity is due to the motor commands generating forces that engage the environment in different ways. Reexecuted in a different task, these motor commands are *a priori* more likely to generate a diverse set of effects – and thus information – than a set of motor commands that produced the same effect in the source task.

We can interpret the method in the context of the *learner-as-a-scientist* paradigm; it can be viewed as creating a repertoire of experiments to conduct in unknown situations to discover how the environment behaves and what interactions it responds to. This type of behavior is seen in nature: "A young corvide bird, confronted with an object it has never seen, runs through practically all of its behavioral patterns, except social and sexual ones" (Lorenz, 1996).

Likewise, a robot interacting with a ball needs to use different movements to make it roll left, right or forward. Having learned
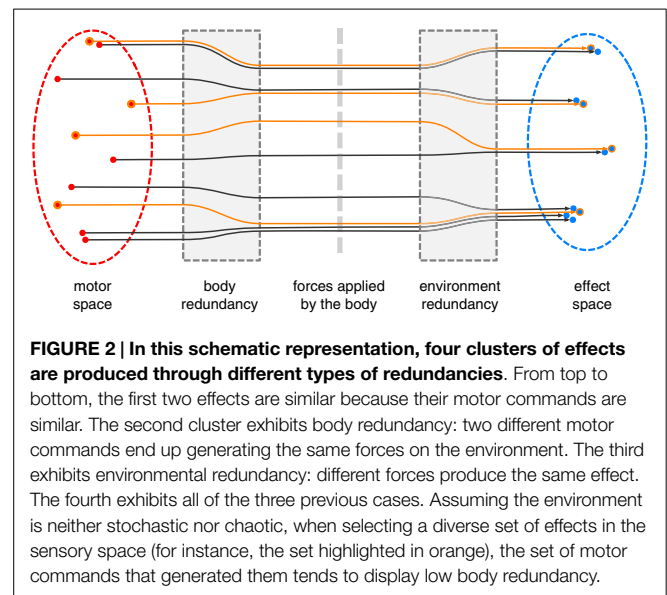
those movements, if the robot is provided with a cube, the prediction or control model of the ball is difficult to exploit directly: the two objects have significantly different dynamics. However, by reusing the behavioral patterns – the movements – on the cube that pushed the ball in different directions, the robot can immediately produce a diversity of effects on the cube, and start learning which ones still apply and are most effective.

Selecting motor commands through diversity can be understood as trying to filter body redundancy. In a given task, the redundancy of the body and the environment will make different motor commands produce the same effect, as **Figure 2** illustrates. If the body redundancy is at play, two different motor commands will end up applying similar forces on the environment: this is the case of a redundant multi-joint robotic arm, where multiple motor motions exist that generate the same end-effector trajectory. If the environmental redundancy is at play, different forces will produce the same salient effect: this is the case when pushing or pulling on a closed door. A set of motor commands that produce a diversity of effects tends to display neither body nor environmental redundancy. When the environment changes, the absence of body redundancy is conserved among this set. And if the new environment is similar to the old one, some of the environmental redundancy may be avoided as well.

Of course, a stochastic or chaotic environment can counterbalance its redundancy: the same motor command executed multiple times can generate diverse effects. In that case, however, reexecuting this motor command multiple times in the new task is a justified strategy to generate diversity.

In the following, we detail first how the source task is explored, and the learning algorithms we use. Then, we explain how the exploration is modified for the target task.

3.1. Exploration of the Source Task

In this paper, the source task will be explored using a goal-directed exploration algorithm. Goal-directed exploration (Oudeyer and Kaplan, 2007) implementations have been

proposed in Baranes and Oudeyer (2010), Jamone et al. (2011), and Rolf et al. (2011), and as part of the SAGG-RIAC architecture (Baranes and Oudeyer, 2013) and have been shown to be effective in exploring sensorimotor spaces with large motor spaces. These methods for goal babbling as well as related methods such as MAP-Elite (Cully et al., 2015) have also been shown to efficiently generate forms of behavioral diversity.

In what follows, we introduce and use the EXPLORE algorithm, a variant of the SAGG-Random goal babbling algorithmic architecture (Baranes and Oudeyer, 2013; Moulin-Frier and Oudeyer, 2013). We adapt SAGG-Random by adding a bootstrapping phase of random motor babbling lasting K_{boot} steps before the random goal babbling phase (**Algorithm 1**); the bootstrapping phase is necessary because the inverse models we use need some existing data to work. During the bootstrapping phase, random motor commands are executed, while during the random goal babbling phase, random goals are chosen uniformly in S, and an *inverse model* (introduced in Section 3.2) is used to transform goals into motor commands. In the experiments, we set K_{boot} to a low value in order to reduce the duration of the random motor babbling phase, without significantly compromising performance. In a more general context, K_{boot} could be computed dynamically, for instance, by using the method introduced in Benureau and Oudeyer (2015).

We use here two implementations of this architecture, corresponding to two different learning algorithms, INVERSEPERTURB and INVERSELBFGSB-LWLR, to implement the INVERSE step, as described in the next section.

3.2. Inverse Model

An inverse model is used whenever goal babbling is chosen as an exploration strategy in the EXPLORE algorithm. Let us remark that our objective here is *not* to acquire a forward or inverse model of the environment. The learning algorithms are functional entities of the exploration process, and the models they produce are not evaluated. In particular, they may make assumptions that preclude them from creating accurate models of the environment – we will discuss such a case in Section 4. In this article, we will be using two different inverse models: a simple, perturbation-based one and another based on an optimized regression method.

3.2.1. Perturbation-Based Inverse Model

The perturbation-based model finds the best motor command to reach the goal among those already executed in the past and creates a slightly perturbed variation of it to be executed, in a fashion similar to the mutation operators of evolutionary algorithms.

Given a motor command $\mathbf{x} = \{x_0, x_1, \ldots, x_{d_M-1}\}$ in M, a perturbation of $\mathbf{x}$ is defined by:

$$\text{P}_{\text{ERTURB}_d}(\mathbf{x}) = \{\text{R}_{\text{ANDOM}}(\max(a_i, x_i - d(b_i - a_i)),$$
$$\min(x_i + d(b_i - a_i), b_i))\}_{0 \le i < d_M}$$

with $M = \prod_{0 \le i < d_M} [a_i, b_i]$ as a hyperrectangle of $\mathbb{R}^{d_M}$ and with the function $\text{R}_{\text{ANDOM}}(a,b)$ drawing a random value in the interval [a,b] according to a uniform distribution. d is the perturbation parameter, and belongs to [0, 1]; it is the only parameter of the inverse model, that we can now express in **Algorithm 2**.

ALGORITHM 1 | EXPLORE $((f, n), K_{boot})$.

Input: (f, n)	▷ Exploration task.
Input: K_{boot}	▷ Duration of random motor babbling.
Output: $H = \{\mathbf{x}_j, \mathbf{y}_j\}_{0 \le j < n} \in (S \times M)^n$	▷ Exploration history.
$H \leftarrow \varnothing$	
for $t: 0 \rightarrow n-1$ **do**	
if $t < K_{boot}$ **then**	▷ Run motor babbling for the first K_{boot}
$\mathbf{x}_t \leftarrow \text{MOTORBABBLING}(M)$	steps.
else	▷ Then switch to goal babbling.
$\mathbf{x}_t \leftarrow \text{GOALBABBLING}(S, H)$	
$\mathbf{y}_t \leftarrow f(\mathbf{x}_t)$	▷ Execute the motor command.
add $(\mathbf{x}_t, \mathbf{y}_t)$ to H	▷ Update the history.
procedure MOTORBABBLING(M)	▷ Motor babbling does not depend on
choose $\mathbf{x}_t$ randomly in M	history.
return $\mathbf{x}_t$	
procedure GOALBABBLING(S, H)	▷ Corresponds to the SAGG-Random
choose a goal $\mathbf{g}_t$ randomly in S	architecture.
$\mathbf{x}_t \leftarrow \text{INVERSE}(\mathbf{g}_t, H)$	▷ Use the inverse model on the goal to
return $\mathbf{x}_t$	produce a motor command.

ALGORITHM 2 | INVERSEPERTURB$_d(g_t, H)$.

Input: $\mathbf{g}_t \in S$	▷ A goal.
Input: $H = \{(\mathbf{x}_j, \mathbf{y}_j)\}_{0 \le j < t-1} \in (M \times S)^{t-1}$	▷ Current history: past $t-1$ observations.
Input: $d \in [0, 1]$	▷ Perturbation ratio.
Output: $\mathbf{x}_t \in M$	▷ Candidate motor command.
$\mathbf{x}_{nn}, \mathbf{y}_{nn} \leftarrow \text{argmin}_{x_j, y_j \in H} (\|\mathbf{y}_j - \mathbf{g}\|)$	▷ $\mathbf{y}_{nn}$ is the nearest neighbor of $\mathbf{g}$.
$\mathbf{x}_t \leftarrow \text{Perturb}_d(\mathbf{x}_{nn})$	▷ Perturbing the motor command of the nearest effect.

ALGORITHM 3 | INVERSELBFGSB-LWLR(g_t, H).

Input: $\mathbf{g}_t \in S$	▷ A goal.
Input: $H = \{(\mathbf{x}_j, \mathbf{y}_j)\}_{0 \le j < t-1} \in (M \times S)^{t-1}$	▷ Current history: past $t-1$ observations.
Output: $\mathbf{x}_t \in M$	▷ Candidate motor command.
$\mathbf{x}_{nn}, \mathbf{y}_{nn} \leftarrow \text{argmin}_{x_j, y_j \in H} (\|\mathbf{g}_t - \mathbf{y}_j\|)$	▷ $\mathbf{y}_{nn}$ is the nearest neighbor of $\mathbf{g}_t$.
Initialize L-BFGS-B optimization with $\mathbf{x}_{nn}$	
$\mathbf{x}_t \leftarrow \text{MinimizeLBFGSB}_{\mathbf{x} \in M}(\|\mathbf{g}_t - \text{PredictLWLR}(\mathbf{x}, H)\|)$	

This perturbation-based inverse algorithm is simple and effective. Its complexity is linear in both d_M (perturbation) and nd_s (nearest neighbor search). Its main assumption is that a small perturbation of the motor space produces a comparatively small change in the sensory feedback. The model has difficulties escaping local minima. In practice, in the experimental contexts considered in this paper, the performance and robustness of this model is competitive with more complex approaches. This inverse model is not completely unreasonable in biological organisms (Loeb, 2012), and that related algorithms implemented as part of the SAGG-Random architecture such as in Baranes and Oudeyer (2013) and Moulin-Frier et al. (2014), as well as other variations such as the MAP-Elite algorithm (Cully et al., 2015), have yielded good results in diverse robotics contexts.

3.2.2. Optimized Regression Inverse Model

We also use an optimized regression inverse model in some experiments, based on an optimization routine, L-BFGS-B

(Byrd et al., 1995; Zhu et al., 1997), and a predictor, Locally Weighted Linear Regression (LWLR) (Cleveland and Devlin, 1988; Atkeson et al., 1997a,b).

3.2.2.1. Forward Model

To approximate the function f from a set of observations, we employ Locally Weighted Linear Regression (LWLR) (Cleveland and Devlin, 1988; Atkeson et al., 1997a,b), an incremental machine learning algorithm. Although LWLR is more sophisticated than the perturbation-based inverse model, it is still a simple method compared to the state-of-the-art. Here, the absolute learning performance is of little concern as we are interested in comparing different exploration strategies. Still, LWLR is reasonably robust (Munzer et al., 2014) for the learning tasks we are considering. Compared to the perturbation-based inverse model, LWLR is able to extrapolate, i.e., the distance between the goal and the existing observations is taken into account, but it also needs several closely clustered observations to do so efficiently; the perturbation-based inverse model only ever needs one.

Given a set of observations $H = \{(\mathbf{x}_j, \mathbf{y}_j)\}_{0 \le j < t-1}$ where for each j, $f(\mathbf{x}_j) = \mathbf{y}_j$, and given a query vector $\mathbf{x}_q$, for which we wish to predict the effect, we compute the Euclidean distance to $\mathbf{x}_q$ from each point $\mathbf{x}_j$, and derive the following Gaussian weights w_j:

$$w_j = e^{\frac{-\|\mathbf{x}_j - \mathbf{x}_q\|^2}{2\sigma^2}}$$

We consider the matrices X with $X_{i,j} = (x_i)_j$, Y with $Y_{i,j} = (y_i)_j$, and $W = \mathrm{diag}(w_0, w_1, \ldots w_{t-2})$, and compute:

$$\beta = ((WX)^T WX)^+ ((WX)^T WY)$$

where $(WX)^T WX$ is a symmetric matrix, and $((WX)^T WX)^+$ is its Moore–Penrose inverse (Penrose and Todd, 1955). Then,

$$\mathbf{y}_q = \mathbf{x}_q \beta$$

$\mathbf{y}_q$ is the LWLR estimate of $\mathbf{x}_q$, given the observed data H. We call PREDICTLWLR($\mathbf{x}_q$, H) the function that computes $\mathbf{y}_q$ for any $\mathbf{x}_q \in M$ given H.

In our implementation, σ, which controls the locality of the regression, is dynamically computed. We compute σ as the average distance of the $k = 2d_M + 1$ closest points of the query vector $\mathbf{x}_q$. All other points of H besides the k closest neighbors are given a weight of zero.

3.2.2.2. Inverse Model

Given a goal $\mathbf{g}_t \in S$, we want to produce a motor command $\mathbf{x}_t \in M$ so that $\|f(\mathbf{x}_t) - \mathbf{g}_t\|$ is small.

With M being a hyperrectangle of $\mathbb{R}^{d_M}$, we use L-BFGS-B (Limited-memory Broyden–Fletcher–Goldfarb–Shanno Bound constrained (Byrd et al., 1995; Zhu et al., 1997), version 3.0 (Morales and Nocedal, 2011)), a quasi-Newton method for bound-constrained optimization, to minimize the error. L-BFGS-B use an approximation of the Hessian matrix to direct the optimization (because the Hessian cannot be directly computed, it is approximated using finite differences). We approximate $\|f(\mathbf{x}) - \mathbf{g}\|$ with $\|\mathrm{PREDICTLWLR}(\mathbf{x}, H) - \mathbf{g}\|$ and use it with L-BFGS-B to further approximate $\mathrm{argmin}_{x \in M}(\|f(\mathbf{x}) - \mathbf{g}\|)$.

The optimization process is initialized with the motor command corresponding to the closest neighbor of $\mathbf{g}$ in the set of observations (see **Algorithm 3**).

The INVERSE method is replaced by either INVERSEPERTURB or INVERSELBFGSB-LWLR in the source task exploration algorithm, EXPLORE, and the one of the target task, REUSE, that we introduce now.

3.3. Exploration of the Target Task

The exploration of the target task is organized around two algorithms. The first, TRANSFER, is applied at the end of the interaction with the source task and produce a set of motor commands bins that are used by the second, REUSE, to affect the exploration of the target task.

The TRANSFER selects motor commands that produced a diversity of effects. It works by partitioning the sensory space of the source task, S_A. We use a simple grid here. To each cell of the grid corresponds a bin of motor commands that contains all the motor commands whose effects belong to the cell. This way, similar effects in the source task have their motor command gathered in the same bin (see **Algorithm 4**).

The REUSE method is a variation of the EXPLORE algorithm, where a part of the random motor babbling steps are replaced by *reuse steps*. During a reuse step, a random bin among the ones generated by the TRANSFER algorithm is selected, and a random motor command is drawn from the bin without replacement and executed in the environment. Such a selection generates a sequence of motor commands that correspond to effects representative, on average, of the diversity of effects produced in the source task. Goal babbling behavior is unaffected.

To produce the REUSE method, the call to MOTORBABBLING in the EXPLORE algorithm is replaced by a probabilistic call to REUSEBABBLING and MOTORBABBLING, according to a probability p_{reuse} (see **Algorithm 5**).

This procedure has a low computational cost, and only transfers structured sets of motor commands between tasks. No sensory data are shared across tasks, which mean that no forward or inverse model is shared. It makes the method compatible with arbitrary changes in sensory modalities, and insensitive to the quality forward or inverse models of the source task, should they exist. Furthermore, by separating the TRANSFER and REUSE method, we can precompute the transferred data before the second task is known, and then use it even if the sensory data of the first task has been forgotten.

Here, we have proposed a TRANSFER method that partitions the sensory space of the source task. This partitioning encodes

---

ALGORITHM 4 | TRANSFER(H_A).

Input: $H_A = \{(\mathbf{x}_j, \mathbf{y}_j)\}_{0 \le j < n_A}$ ▷ *Exploration history of the source task.*
Input: A partitioning method.   ▷ *We use grid partitioning in this paper.*
Output: $\mathcal{B}$ ▷ *Set of motor commands bins.*

Partition S_A according to the partitioning method, and to each region, assign a bin.
The set of all bins is $\mathcal{B}$
for $(\mathbf{x}_j, \mathbf{y}_j) \in H_A$ **do**
     add $\mathbf{x}_j$ to $\mathrm{bin}_{\mathcal{B}}(\mathbf{y}_j)$ ▷ $\mathrm{bin}_{\mathcal{B}}(\mathbf{y}_j)$ *is the bin of the region of S_A*
         where $\mathbf{y}_j$ belongs

ALGORITHM 5 | REUSE((f_B, n_B), $\mathcal{B}$, K_{boot}, p_{reuse}).

Input: (f_B, n_B)	▷ *Target task.*
Input: $\mathcal{B}$	▷ *Set of bins of motor commands.*
Input: K_{boot}	▷ *Duration of bootstrapping.*
Input: p_{reuse}	▷ *Ratio of transfer motor babbling.*
Output: $H_B = \{(\mathbf{x}_j, \mathbf{y}_j)\}_{0 \le j < n_B} \in (S_B \times M_B)^{n_B}$	▷ *Observation history.*

   $H_B \leftarrow \varnothing$
   for $t: 0 \to n_B - 1$ **do**
      if $t \le K_{boot}$ **then**
         if RANDOM$(0, 1) \le p_{reuse}$ **then**   ▷ *Probabilistic call to reuse.*
            $\mathbf{x}_t \leftarrow$ REUSEBABBLING$(M_B, \mathcal{B})$
         else
            $\mathbf{x}_t \leftarrow$ MOTORBABBLING(M_B)
      else
         $\mathbf{x}_t \leftarrow$ GOALBABBLING(S_B, H_B)
      $\mathbf{y}_t \leftarrow f_B(\mathbf{x}_t)$ ▷ *Execute the command.*
      add $(\mathbf{x}_t, \mathbf{y}_t)$ to H_B ▷ *Update the history.*
   Procedure REUSEBABBLING$(M_B, \mathcal{B})$
      if at least one bin of $\mathcal{B}$ is not empty **then**
         choose a non-empty bin b_R of $\mathcal{B}$ randomly.
         draw $\mathbf{x}_t$ from b_R without replacement
         return $\mathbf{x}_t$
      else   ▷ *Revert to motor babbling*
         return MOTORBABBLING(M_B) *if no reusable command is available.*

diversity, and may be non-trivial in complex sensory spaces. There is flexibility in how the TRANSFER method could be implemented however. It could, through an arbitrary method – optimization of a diversity measure for instance – build a small set of motor commands whose effects have high diversity, and return a single bin containing them to the REUSE method, discarding other observations from the source task. The REUSE method would select randomly from this single bin as a result.

In the following sections, we conduct experiments to show that REUSE is effective in situations that involve changes in the morphology of the robot (arms with different link lengths in Section 4), that involve switching an object for another between the source and the target task (ball/cube experiment in Section 5.2), exploiting pure random motor babbling (Section 5.2.2), dealing with dissimilar situations (Section 5.2.3), and scaffolding ones (pool experiment in Section 5.3). We also investigate how REUSE can be used to exploit simulation results on real robots (Section 5.5).

4. EXPERIMENT ON PLANAR ARMS

To illustrate the REUSE method, let us consider a pair of planar robotic arms, each with 20 joints. The first arm has same-length links totaling one meter, and the environment returns the Cartesian position of the end-effector. The second arm has links such that, going from the base to the end-effector, each link is 0.9 times smaller than the previous one, while the total length of the arm remains one meter; this arm also returns the position of the end-effector, but using *polar* coordinates (**Figure 3**).[1]

---

[1]The source code and data for producing all graphs is published (Benureau and Oudeyer, 2016) and is made available at https://dx.doi.org/10.6084/m9.figshare.2816284

FIGURE 3 | When executing the same command on both arms, the position of the end-effector is significantly different most of the time. Here depicted are 50 pairs of executions of the same motor command on the two 20-joint arms, five of which that are highlighted.
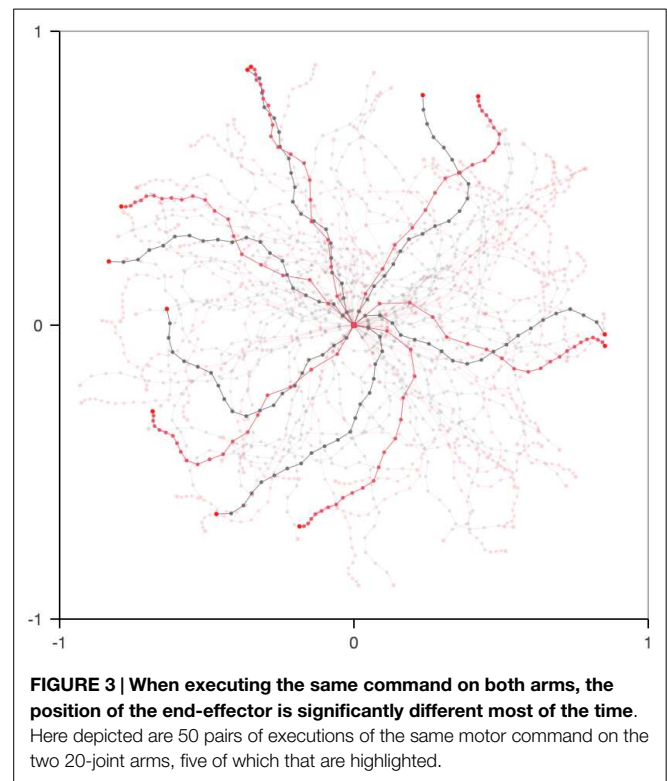
The two arms have a different morphology – a situation akin to morphological development. They share, however, the same number of joints with the same available ranges ($\pm 150°$): they have the same motor space and motor parameterization. However, because the lengths of the links are different, most motor commands will result in a different position for the end-effector, as shown in **Figure 3**. And because the positions are expressed in two different coordinate systems, the inverse model of one arm is difficult to exploit on the other arm, without having, or learning, a mapping between the coordinate systems.

The exploration on the first arm is conducted over 5000 steps, using the EXPLORE algorithm with $K_{boot} = 50$, with the perturbation-based inverse model with $d = 0.05$, i.e., perturbing each joint by at most $\pm 15°$.

The exploration of the target arm is the same as the source arm, except that all the 50 motor babbling steps of the source exploration strategy are replaced by reuse steps, as per the REUSE algorithm with $p_{reuse} = 1$. **Figure 4A** illustrates how motor commands to be reused are selected, as per the TRANSFER algorithm. **Figures 4B,C** show the difference between the bootstrapping phase of the REUSE and EXPLORE algorithm. The impact of REUSE on the exploration is important at the beginning and remains beneficial throughout, even after 5000 steps.

Figure 5 displays the τ-coverage (with $\tau = 0.05$) of both the REUSE and the EXPLORE algorithm on the target arm over 100 repetitions of the experiment. In both cases, the coverage was computed in the Euclidean space. The REUSE strategy provides a performance increase that last even after 5000 steps: in 75% of the cases, the REUSE strategy performs strictly

FIGURE 4 | Illustration of the REUSE algorithm. After the end of the exploration of the source arm **(A)**, a 20×20 grid partitions the effect space, and as many times as necessary (50 times here), a random cell is chosen, as well as a random effect inside it (dots highlighted in red). The motor commands that produced the chosen effects are then reexecuted on the target arm **(B)**. This replaces the initial 50 motor babbling steps of the EXPLORE algorithm **(C)**. In both cases, the effects produced by random motor babbling or a reused motor command have been highlighted in red. While random motor babbling produces convoluted arm postures whose effects are clustered around the center, the reused motor commands produce effects spread out over the reachable space, and feature straighter postures. This difference in bootstrapping has a huge impact on the coverage at $t = 400$, and a lesser, but still present one after 5000 steps.

better than the best-case scenario of the EXPLORE strategy. The usage of REUSE accelerates the exploration of the reachable space.

However, an interesting phenomenon is present. The worst case of the REUSE strategy, as shown by the dotted lines, performs worse than the worst case of the EXPLORE strategy.

FIGURE 5 | The exploration with Reuse on the second arm covers significantly more of the reachable space early on than the one without. The graph shows the τ-coverage (with $\tau = 0.05$) of the second arm explored with the REUSE and EXPLORE algorithm respectively, over 100 repetitions of each experiment. The median case is displayed, surrounded by a margin going from the 25th to the 75th percentiles. In dashed lines, the worst and the best coverages are also pictured.

To understand why, it is interesting to look at goal babbling as an evolutionary algorithm. From an evolutionary robotics perspective, the motor commands are the genetic encoding, the arm posture the phenotype and the effect – the position of the end effector – is the behavior of the arm. At each timestep of goal babbling, a random goal is chosen. The distance to this goal defines a fitness function, and the highest-performing past observation, whose effect is the nearest neighbor of the goal, is chosen to reproduce through mutation: this is how our perturbation-based inverse model works.

Therefore, after the bootstrapping phase, arm postures are chosen for reproduction in proportion of how close their effects are to the chosen goals. When using random motor babbling, most postures produce effects near the center. Because goals are randomly chosen in the $[-1,1] \times [-1,1]$ square, most goals are farther from the center than most observed effects. It means that postures producing effects on the edge of the initial cluster are chosen and mutated with disproportionate frequency. Through repeated selection and mutations those postures and their descendants, straighter and straighter postures are discovered.

Sometimes, however, those initial arm postures contain loops. Those loops represent local minima that are difficult to escape. The mutation and selection process – our perturbation-based inverse model – tends to straighten arm postures to reach distant target. In the process, loops are tightened, not removed. Therefore, the maximum span of the arm is reduced, and the exploration covers only a fraction of the reachable space, as shown in the graphs of **Figure 6**.

On the source arm, because the links are all of the same length, loops have the same cost in span regardless of where they appear. But on the target arm, they are most costly near the base of the arm, where links are longer. Therefore, arm postures featuring loops near the base of the arm tend to be shorter on average than postures with loops near the tip, even in a random motor babbling sampling. It means that, when using the EXPLORE algorithm, most of the time, those postures do not get selected for far goals after the bootstrapping phase, as better solutions exist. Therefore, the postures that explore the edge of the reachable space have a

tendency to have either loops near the tip of the arm or no loops at all.

The only way for postures with costly loops near the base of the arm to be selected on the target arm is for them to have the rest of the arm rather straight, and in a fashion disproportionate with the other arm postures they compete with. This is exactly the scenario that happens in the worst case of REUSE: all the reused arm postures where the tip is far from the center have loops near the base of the arm, as **Figure 6** illustrates. This is not a problem for the source arm, but for the target arm it limits the achievable span much more than if the loops were near the tip of the arm.

This explains the difference in coverage between the worst case of the REUSE and EXPLORE algorithm on the target arm, and serves to illustrate a danger of the REUSE algorithm: providing good solutions tra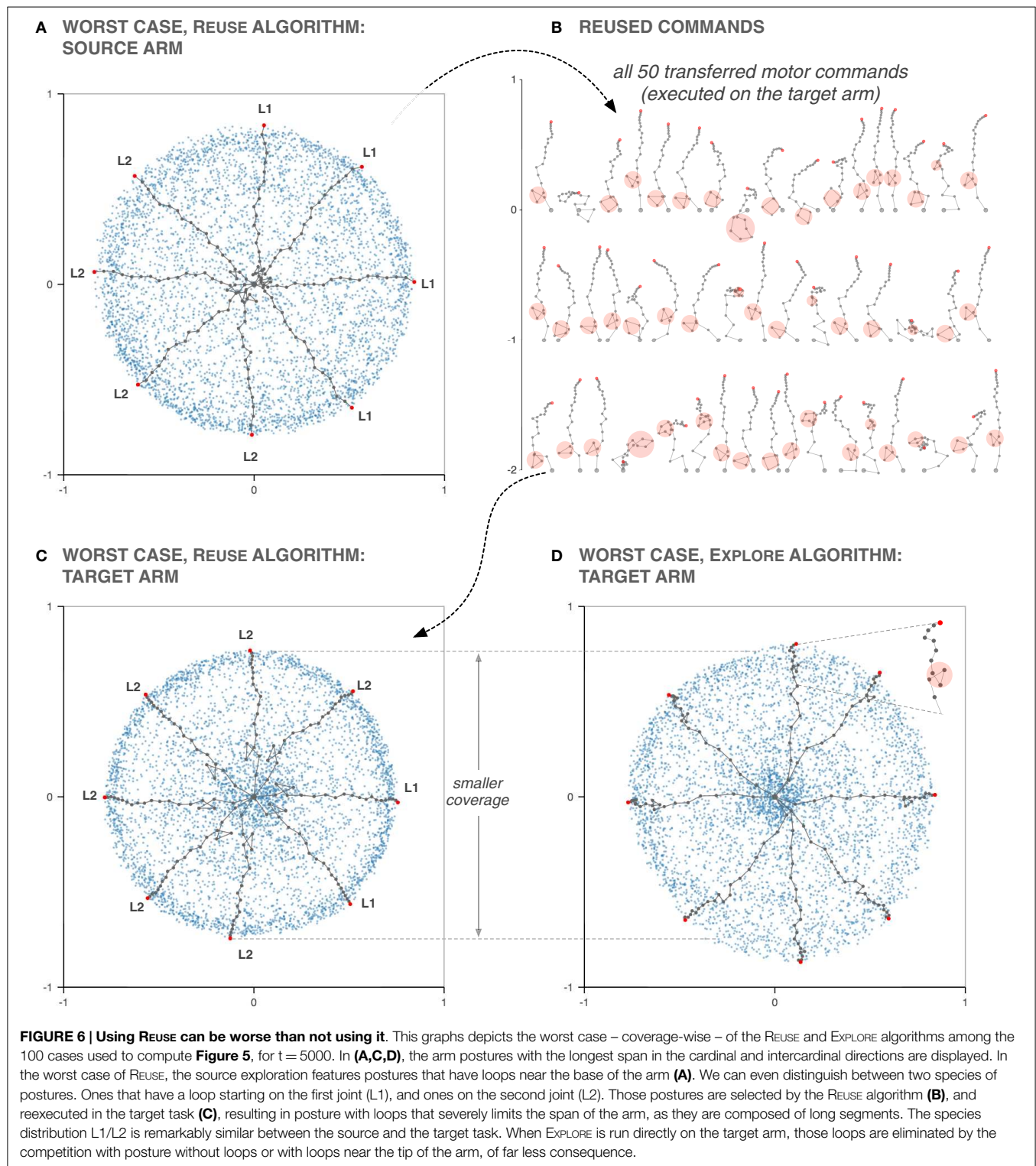pped in local minima early in exploration can prevent the discovery of better solutions, more adapted to the target task. Let us remark here that all the 50 motor babbling steps of the EXPLORE algorithm were replaced by reuse steps in the REUSE algorithm. But allowing a portion of the 50 steps to remain random motor babbling, for instance with $p_{reuse} = 0.5$, would not solve the problem (we tested), as the arm postures with the best span in the bootstrapping phase would remain the reused ones, and get selected and mutated more than the others.

Of course, the occurrence of such a problem is highly contingent on the specifics of the two tasks, on how goals are chosen and what inverse model is used. But the risk, when transferring knowledge or skill from one task to another, to negatively impact the performance in the target task is always a possibility, and is difficult to protect from inside the framework of the problem we are considering.

Still, this does not mean that REUSE should be avoided. While it has the potential to induce performance-hindering local minima, it also has the potential to propose good solutions early in exploration. During the first 150 steps, the worst case scenario of the REUSE algorithm is actually better than the best-case scenario of the EXPLORE one. In a robotic and operational context, having good-enough solutions quickly might matter more than finding perfect ones eventually. Robots do not live at the asymptote. If a robot needs to learn how to whisk for a recipe, it may matter more than the eggs and milk are mixed under 15 minutes than the fact that the quickly discovered whisking motion consumes more energy, is less efficient and makes more noise than necessary. Even in a learning context, having good early performance can help decide quickly if the skill is possible to learn, worth learning, and can help form an estimation of what is achievable in the target task, which may in turn quickly bootstrap planning capabilities.

Before moving on to a more complex experimental setup, it is interesting to analyze why the REUSE method is effective. As pointed out before, the two arms have different inverse models, and the relationship between them is non-trivial. By reusing motor commands that produce a diversity of effects, we make the assumption that the diversity mapping is simpler between the two tasks: a set of motor commands generating a certain amount of diversity on the source arm will generate a similar amount on the target arm. This is something we can verify experimentally.

FIGURE 6 | Using REUSE can be worse than not using it. This graphs depicts the worst case – coverage-wise – of the REUSE and EXPLORE algorithms among the 100 cases used to compute **Figure 5**, for t = 5000. In **(A,C,D)**, the arm postures with the longest span in the cardinal and intercardinal directions are displayed. In the worst case of REUSE, the source exploration features postures that have loops near the base of the arm **(A)**. We can even distinguish between two species of postures. Ones that have a loop starting on the first joint (L1), and ones on the second joint (L2). Those postures are selected by the REUSE algorithm **(B)**, and reexecuted in the target task **(C)**, resulting in posture with loops that severely limits the span of the arm, as they are composed of long segments. The species distribution L1/L2 is remarkably similar between the source and the target task. When EXPLORE is run directly on the target arm, those loops are eliminated by the competition with posture without loops or with loops near the tip of the arm, of far less consequence.

In **Figure 7**, the coverage of sets of random motor commands of diverse sizes is highly correlated between the two arms. This correlation in the production of diversity is therefore an assumption it seems possible to rely on and exploit, even, in some cases, when the sensory modalities or the morphology of the robots are different between tasks.

5. OBJECT INTERACTION TASKS

5.1. Experimental Setup

We consider an experiment where a robotic arm interacts with an object and observes its displacement at the end of the interaction. In a developmental context, an interaction task is relevant, as it

FIGURE 7 | The diversity production of the same-length-links arm and the increasing-length-links arm is highly correlated. The graph displays the coverage of 1000 different sets of random motor commands of size 1, 2, . . ., 1000, respectively, for the 20-joint source and target arm, with $\tau = 0.05$.
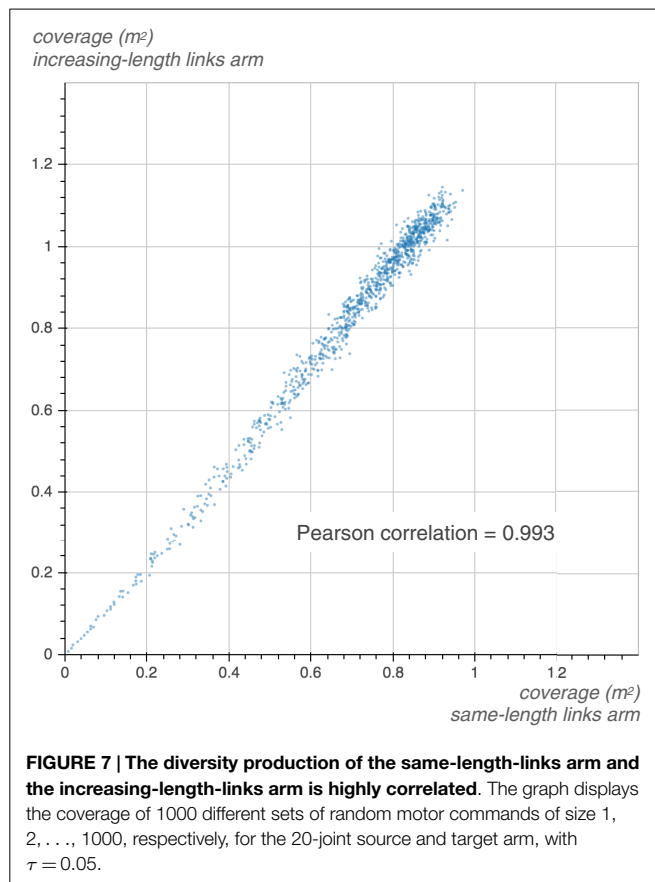
pertains to the early exploration of the world, where the function of most objects is still unknown.

We used both a simulated and a hardware setup, but comparatively few experiments were conducted on the hardware. For this reason, in this section, we mainly focus on describing the simulated setup, but discuss aspects related to the morphology of the real robot as well. The hardware setup is thoroughly described in Section 5.4.

The robot is a serial chain of six servomotors. The three proximal motors are Dynamixel RX-64 and the three distal ones are RX-28. Those servomotors are capable of delivering respectively 6.0 and 2.5 N·m of stall torque, with an angular resolution of 0.29°, measured with a mechanical potentiometer, whose precision is variable (across the angle range and between different motors). During the experiments, the real servomotors were operated in position control mode using the embedded PIDs, with a control loop for the position running at 100 Hz. In simulation, the physical characteristics of the motors are reproduced as much as possible, and their control in position is done in lockstep with the physics engine simulation steps, at 50 Hz.

5.1.1. Dynamic Movement Primitives

The movements of the robot are generated using dynamic movement primitives (DMP). DMPs are parametrized dynamical systems introduced by Ijspeert et al. (2002). They are computed from sets of differential equations that produce smooth movements robust to perturbations. We chose DMPs, and the specific

parameterization we explain below, because it allowed to express many different arm trajectories with a compact description (i.e., few motor dimensions). We use the implementation of Stulp (2014), based on Ijspeert et al. (2013) with the sigmoid variation of Kulvicius et al. (2012).

DMPs are based on damped spring dynamics, perturbed by a forcing term [equation (1)]. The forcing term is a linear combination of basis functions [equation (4)]. Here, Gaussian activation functions $\psi_i(s_t)$ are used, with center c_i and width σ_i, weighted by w_i [equation (3)]. v_t is the phase of the forcing term, described by an sigmoid decay term [equation (2)]. In the following equations, T is the duration of the movement, Δ_t is the time resolution, α, β, and γ are constants and g is the target state.

$$\ddot{x}_t = \alpha(\beta(g - x_t) - \dot{x}_t) + f_t \tag{1}$$

$$\dot{v}_t = -\frac{\gamma e^{\frac{\gamma}{\Delta_t}(T-t)}}{\left(1 - e^{\frac{\gamma}{\Delta_t}(T-t)}\right)^2} \tag{2}$$

$$\psi_i(t) = e^{-\left(\frac{t}{T} - c_i\right)^2 / 2\sigma_i^2} \tag{3}$$

$$f_t = \frac{\sum_{i=0}^{N} \psi_i(t) w_i}{\sum_{i=0}^{N} \psi_i(t)} v_t \tag{4}$$
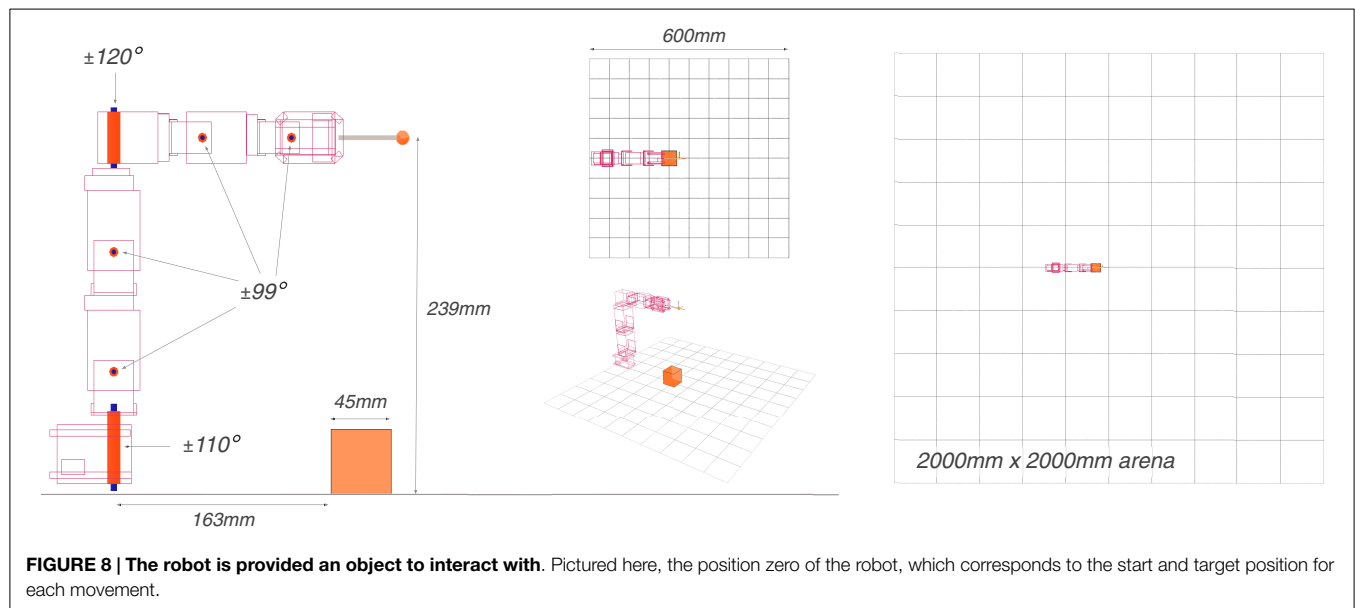
In this experimental setup, the start- and end-points are made identical ($g = x_0$) and correspond to the motor being in the zero position (**Figure 8**). We use two basis functions per motor, with c_0 and c_1 fixed, respectively, at $1/3T$ and $2/3T$, with $T = 2.5$ s (Δ_t is 20 ms and the simulation is stopped at 5 s). σ_0 and σ_1 are shared by all motors.

We do not directly use the weights for parametrizing the motor space. Instead, we use the LWLR function approximator provided with the DMP library (Stulp, 2014), and define two linear functions per motor, with slopes a_0, a_1 and offsets b_0, b_1, respectively. The function approximator then computes the forcing term to approximate as much as possible these functions at time c_0 and c_1. Although directly manipulating the weights would be more natural, this method provides a rich diversity of trajectories, and, because DMPs were not a focus of our work, we did not inquire further about making the system perform better or making the representation more compact. Each motor has independent a_0, a_1, b_0, b_1 parameters, and the motors share σ_0, σ_1, while c_0, c_1 are fixed. With six motors, the motion trajectory of the robot is therefore parametrized by a vector of dimension 26. After solving and integrating the dynamical system, we obtain each motor angular position as a function of time.

To avoid the real robot removing (rather brutally) their own wires, the range of the first and fourth motor from the base are restricted to ±110° and ±120° (**Figure 8**). All other motors are physically restricted by their horns to ±99°. In simulation, the robot has the same angle constraints.

The ranges of the DMP parameters are set, so that 95% of the trajectories of a motor would fall in between the angles the motors were able to produce (using an empirical evaluation), and the rest are clipped to legal motor values.

Before executing the motion on the robot, we check for self-collisions, and collisions with the armature of the experiment. If present, the trajectory is truncated and stops just before the

FIGURE 8 | The robot is provided an object to interact with. Pictured here, the position zero of the robot, which corresponds to the start and target position for each movement.

collision to avoid damage on the real robot. The same collision prevention methods are used in simulation, with the exception that the robot can collide freely with the ground.

5.1.2. Environment and Objects

The simulation is conducted using the robot simulator V-REP (Virtual Robot Experiment Platform), with the Open Dynamic Engine (ODE) as a physics engine backend. The environment features an object placed in a cubic arena. The robot arm can interact with the object and the ground.

We consider two sizes for the arena: 600 mm width and 2000 mm width. The larger arena approximates an unbounded environment, while interactions between the object and the walls are frequent in the smaller one. Unless indicated otherwise, we assume the 600-mm arena is used. Two different objects are used: a ball and a cube, of diameter and width both equal to 45 mm.

As a physics engine, ODE has many undesirable and chaotic behaviors that could be overexploited to produce diversity. For instance, movements where the robot pushes from the top of an object toward the ground yield large and significantly different object displacements over repeated executions.

As a preventive measure, we monitor the forces that are applied between the end effector of the robot and the rest of the environment. If at any point a reactive force exceeds 100 N, the simulation is discarded, and the sensory feedback that would be produced by an immobile robot is returned.

5.1.3. Sensory Primitive

At the end of the simulation, the trajectory of the object is processed by sensory primitives that compute the sensory feedback. We consider a simple sensory primitive that returns the displacement of the object projected on the ground at the end of the simulation. The displacement is returned as a vector of length 3: the displacement in x, in y, and a discrete dimension of saliency, which has value 0 if no collision happened, and 1 otherwise.

The saliency dimension helps separate observations that create collisions from one that do not. This is not crucial for the

perturbation-based inverse model, but it makes the LWLR-based inverse model more robust.

5.1.4. Behavior of the Setup

The simulation environment does not yield repeatable results. Repeated executions of the same movement can generate significantly different effects, as shown in **Figure 9A**. Indeed, the random seed of the physics engine is not reset when the scene is reset.[2] As ODE uses the current state of the random generator to decide the order with which to resolve the constraints at each step, small variations are introduced that are amplified by the chaotic nature of the interaction with the objects.

In **Figure 9B**, the same motor command is executed on the cube and ball task. The same motions do not generate necessarily similar effects on the objects. Moreover, the interaction with the object can significantly impact the trajectory of the end effector.

We ran experiments on the ball task (because the cube occupies more volume, the ball gives a lower estimate of the collision probability) to decide which number of motor babbling timesteps to use during the experiments. By tallying the number of collisions (not counting those that generate too much force) on a large number of random motor babbling steps (25,000), we estimate the probability to interact with the object during a movement at 2.87% for the cube and 1.81% for the ball. To ensure a high probability that every motor babbling phase had at least one collision, we set the bootstrapping phase to 200 steps (resulting in 99.17 and 97.40% probability of at least one collision for the cube and the ball, respectively).

5.2. Cube and Ball Experiments

In this section, we conduct several experiments with the ball and the cube task. All experiments are conducted in simulation. In all experiments, the coverage measure is computed with the radius,

---

[2]This is an implementation detail of V-REP, and there was no way to change it the version we used.

FIGURE 9 | (A) The physics engine is chaotic. When the same motor command is reexecuted multiple times, the variations in object displacement are significant. **(B)** When reusing a motor command moving with the ball on the cube, the produced displacements can be quite different. Let us remark that the interaction with the object can largely impact the robot's motion.
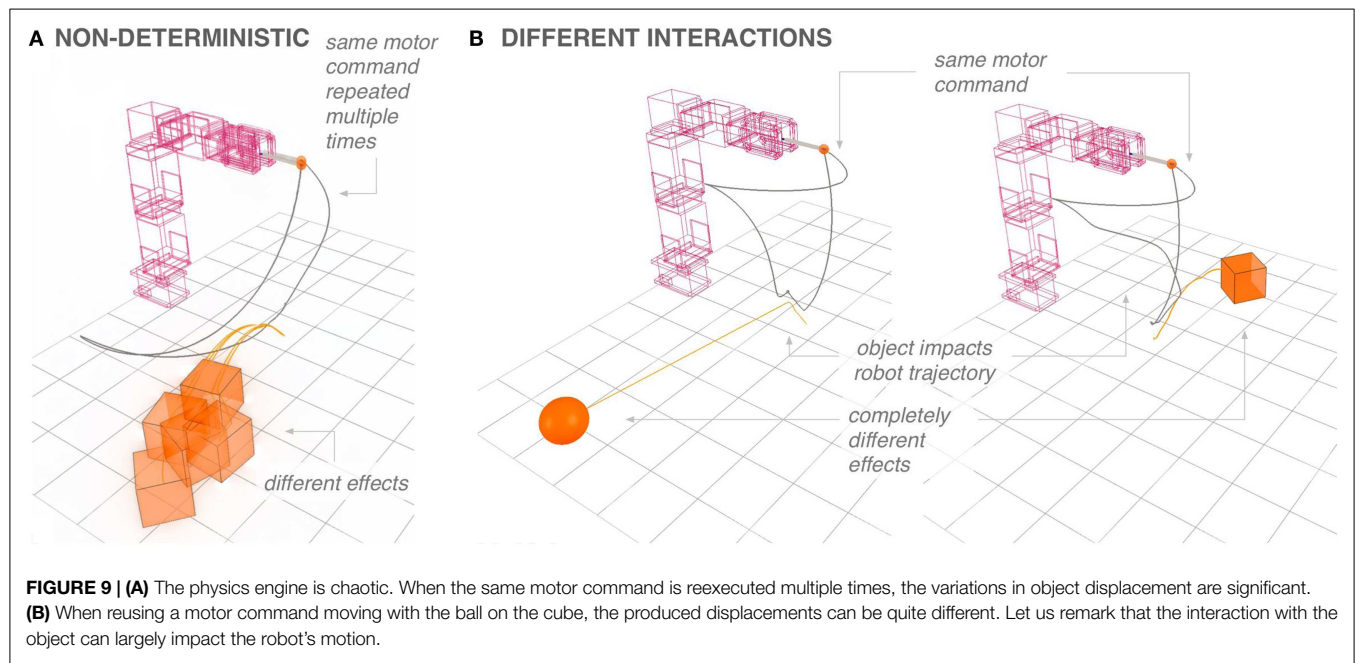
τ, set to 22.5 mm, which is the radius of the ball and the half-width of the cube.

5.2.1. Cubes and Balls

The first experiments look at how REUSE is effective when reusing the exploration of one object for another.

The source task is explored using the EXPLORE algorithm with the perturbation-based inverse model ($d = 0.05$). The random motor babbling phase lasts 200 steps ($K_{boot} = 200$). The target task is explored with the REUSE algorithm, with the same inverse model, and 200 steps of bootstrapping as well. During the bootstrapping phase, each motor babbling step has a 50% probability to be replaced by a reuse steps ($p_{reuse} = 0.5$). In both cases, the exploration lasts 1000 steps in total.

Figure 10 depicts an execution of the REUSE algorithm. The cube is the source task, and the ball is the target task, compared with the ball task using the EXPLORE algorithm. The impact of REUSE is visible during the bootstrapping phase: reusing motor commands from the cube exploration allows to move the ball in many directions in the first 200 steps. In the EXPLORE case, only three interactions are made during that time.

In **Figure 11**, the τ-coverage the four combinations of the cube and ball tasks is shown, for 25 repetitions of the experiment. The REUSE algorithm outperforms the EXPLORE algorithm in all four cases, but the improvements are most important early in exploration. Moreover, when a task uses itself as a source, the impact of the REUSE algorithm is predictably better than when coming from the other object. This is mostly pronounced on the cube task: reusing the ball task is much less effective than when the cube task reuses itself.

A likely explanation of this asymmetry lies in how differently the two objects respond to interaction: the ball will discriminate between most interactions, moving in slightly different directions, while many interactions with the cube will make it just tip over on one side. Therefore, the cube needs more pronounced motions of

the robot to be displaced across the arena, whereas the ball only has to explore small variations of the same movements, which are less effective at generating diversity when reused on the cube.

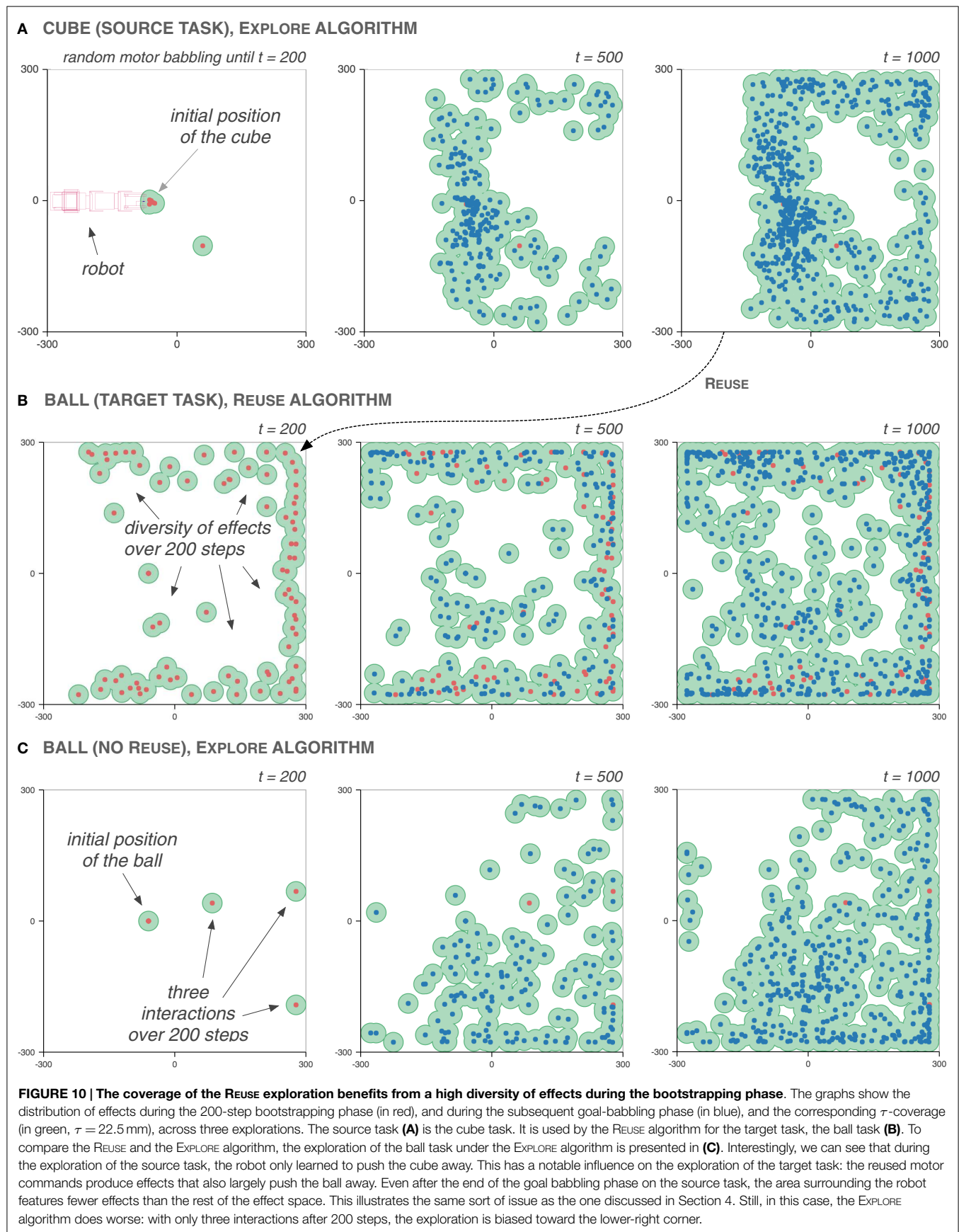### 5.2.2. Different Exploration Algorithms

So far, the source task and the target task have only differed in their exploration algorithm by a few random motor babbling steps replaced by REUSE steps. But the exploration of the source task in not constrained in any such way by the use of the REUSE algorithm.

We consider the case where the source task is explored by a pure random motor babbling strategy. At each of the 1000 steps of the exploration, a random motor command is chosen in the hyperrectangle M and executed. The parameters of the REUSE algorithm remain the same as before. **Figure 12** shows the impact of such a change on the REUSE coverage.

The coverage is improved by reusing motor commands from a random motor babbling source, but less so than when using the EXPLORE algorithm in the source task. The coverage hits a ceiling at around 50 steps into the bootstrapping phase, because the source task did not generate enough diversity to sustain the REUSE algorithm for 200 steps. This leads to the idea of shortening the bootstrapping phase: many times more interactions with the object have been discovered through REUSE after 50 steps than the EXPLORE case will discover through random motor babbling in 200 steps. The goal babbling algorithm has enough observations to be effective.

Figure 13 demonstrates that this is a viable strategy. In the case of the ball task as source task, the coverage improvement in early exploration is actually much greater when the ball task is explored with random motor babbling than with the EXPLORE algorithm (**Figure 11**).

The effectiveness of the REUSE algorithm at exploiting a random motor babbling source also validates the selection process of the motor commands through the diversity of the effects they produced. Indeed, if REUSE was merely selecting random motor

FIGURE 10 | The coverage of the REUSE exploration benefits from a high diversity of effects during the bootstrapping phase. The graphs show the distribution of effects during the 200-step bootstrapping phase (in red), and during the subsequent goal-babbling phase (in blue), and the corresponding τ-coverage (in green, $\tau = 22.5\,mm$), across three explorations. The source task **(A)** is the cube task. It is used by the REUSE algorithm for the target task, the ball task **(B)**. To compare the REUSE and the EXPLORE algorithm, the exploration of the ball task under the EXPLORE algorithm is presented in **(C)**. Interestingly, we can see that during the exploration of the source task, the robot only learned to push the cube away. This has a notable influence on the exploration of the target task: the reused motor commands produce effects that also largely push the ball away. Even after the end of the goal babbling phase on the source task, the area surrounding the robot features fewer effects than the rest of the effect space. This illustrates the same sort of issue as the one discussed in Section 4. Still, in this case, the EXPLORE algorithm does worse: with only three interactions after 200 steps, the exploration is biased toward the lower-right corner.
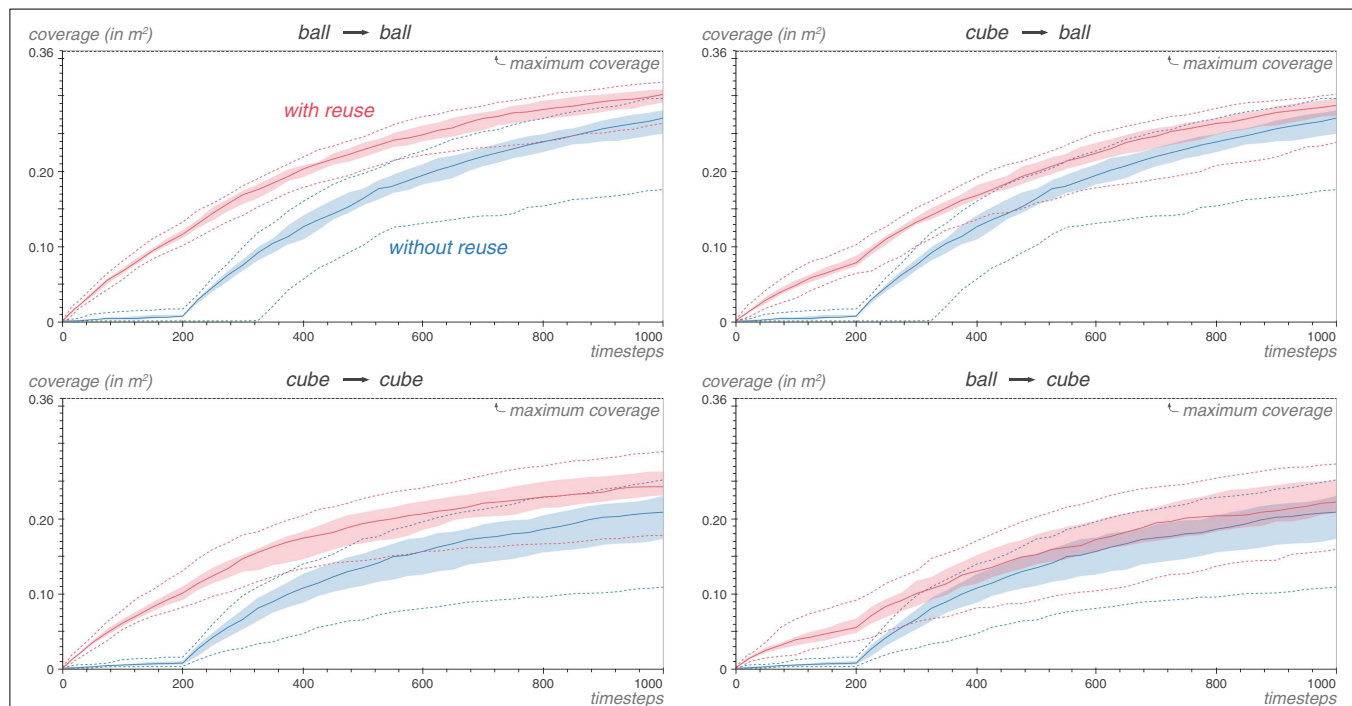
FIGURE 11 | Whether reusing the cube on the ball task or the ball on the cube task, REUSE brings an important coverage boost early in exploration. The figure presents the median of the coverage for the REUSE and EXPLORE cases in the four possible combinations of the cube and ball tasks. The shaded area is delimited by the 25th and 75th percentiles of 25 repetitions of each experiment, and the best and worst case is shown by dashed lines. The effect of REUSE is increased when reusing from the same task, and the ball task is able to exploit the cube task better than the reverse.
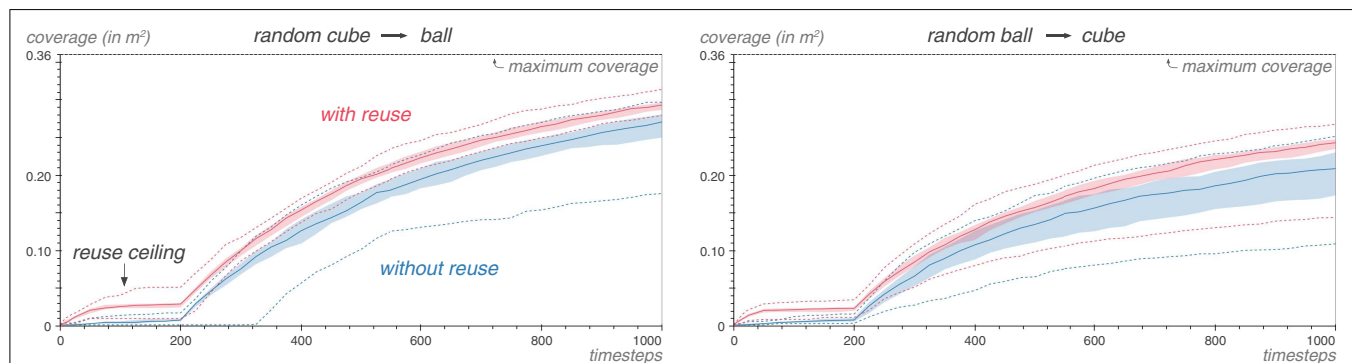

FIGURE 12 | The REUSE method is able to exploit observations generated by random motor babbling. The source task in these graphs was explored using a pure random motor babbling strategy. Repeated 25 times.

commands from the source task, the REUSE method would be equivalent to the random motor babbling strategy when reusing a random motor babbling source: randomly selecting samples from a random source is equivalent to directly sampling the random source. The improvement in coverage here can only be attributed, then, to the selection of motor commands through diversity.
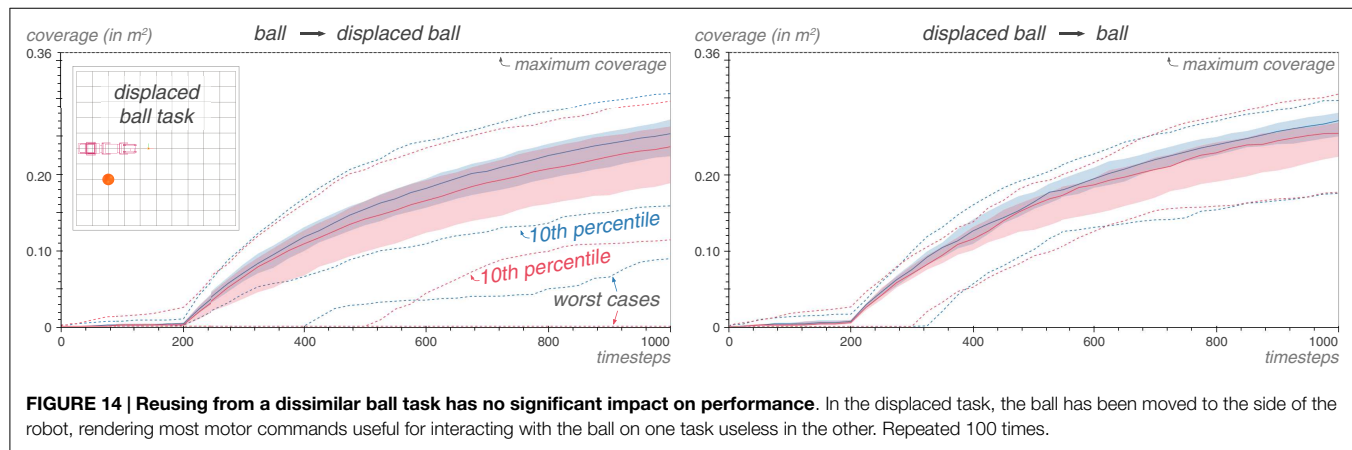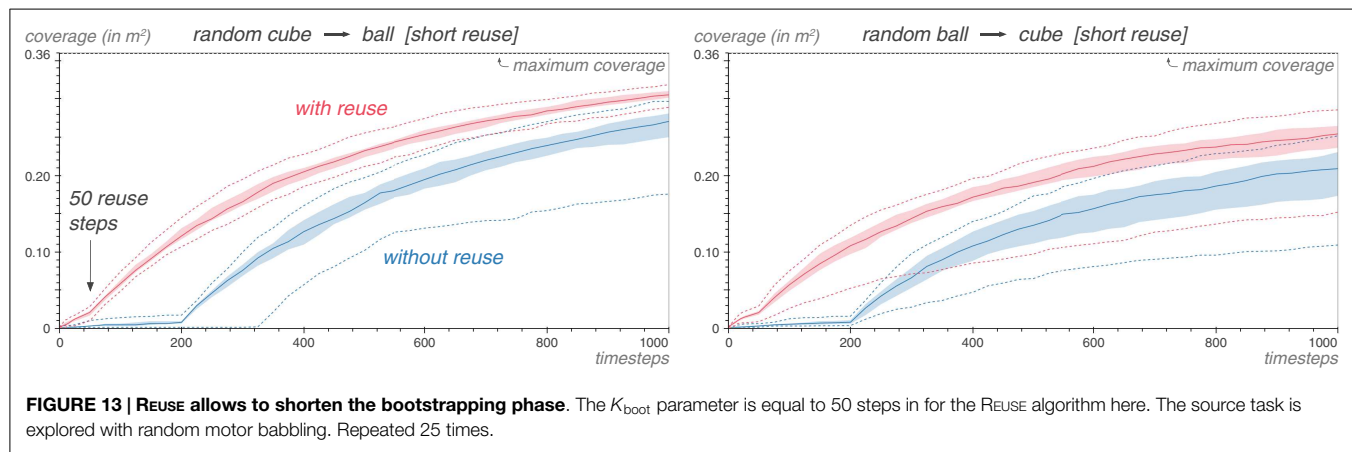
5.2.3. Robustness to Dissimilarity

In the previous experiments, the cube and the ball share the same location relative to the robot. This is of course an important reason for the effectiveness of the REUSE algorithm. While there may be ways for the robot to adapt to such change and still be able to take advantage of REUSE – for instance, by having high-level motor

primitives expressed in an object-centered reference frame – they are not the focus of this article.

However, an important consideration is to examine if the REUSE algorithm can *decrease* the performance of the exploration. The response is of course positive. One can construct a source and a target task so that wasting half of the random motor babbling phase on reusing motor commands guaranteed not to produce any interesting effects could negatively impact the exploration performance. Here, we show that the REUSE algorithm is reasonably robust to a change in the position of the object in the ball environment: some, but not much, of the performance is lost.

In the ball task, the ball is located just under the robot. The *displaced task* is in every way identical to the ball task, except

FIGURE 13 | Reuse allows to shorten the bootstrapping phase. The K_{boot} parameter is equal to 50 steps in for the Reuse algorithm here. The source task is explored with random motor babbling. Repeated 25 times.

FIGURE 14 | Reusing from a dissimilar ball task has no significant impact on performance. In the displaced task, the ball has been moved to the side of the robot, rendering most motor commands useful for interacting with the ball on one task useless in the other. Repeated 100 times.

that the ball has been moved on the right side of the robot. Most movements generating an interaction with a ball will not generate one with the other ball. Moreover, this ball is harder to hit with random movements, with an interaction probability of 0.99% (versus 1.81% before). This is important, because it means that if 100 movements are wasted on reexecuting motor commands that will not hit the ball, the probability of interacting with the ball goes from 86 to 63%.

This is reflected in the results, **Figure 14**. While the loss in coverage at the median is small, the difference is apparent at the 25th percentile in both cases. And the when the displaced task is the target task, the performance below the 25th percentile is much worse for Reuse than for Explore.

While they are not explored here, there are several ways to prevent negative transfer. One is to decrease p_{reuse}, as it decreases the how the Reuse algorithm modifies the original Explore algorithm. When p_{reuse} is equal to zero, both algorithms are equivalent. Another possibility is to dynamically adjust p_{reuse} based on the relative performance of the two bootstrapping strategies: random motor babbling or reused motor commands. We have proposed an algorithmic framework to do precisely that in Benureau and Oudeyer (2015). Ultimately, the decision to use Reuse or not sometimes cannot be made inside the problem we defined: it must come from an external mechanism, which needs only to point out the existence of a relationship between tasks, without specifying
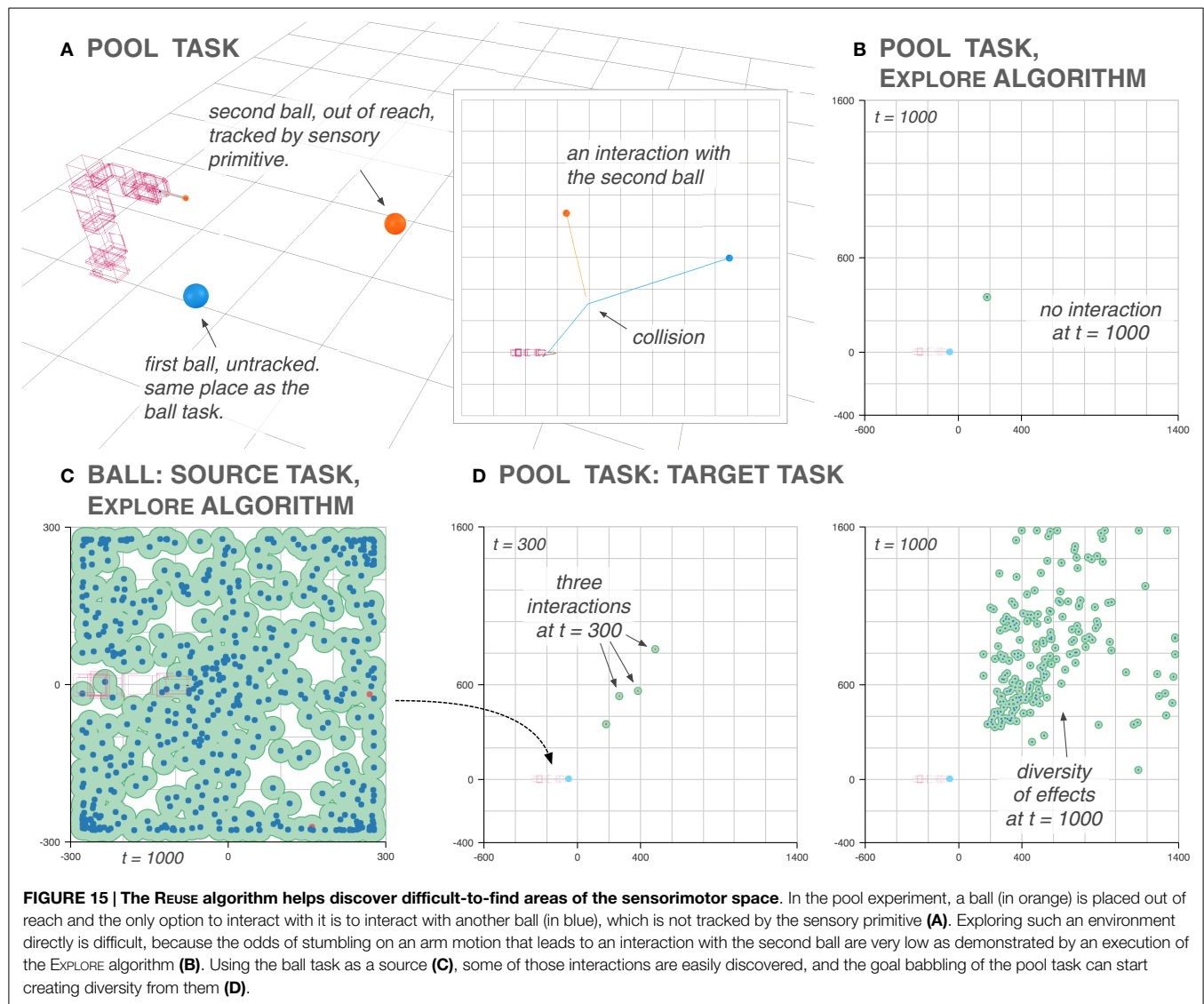
it. We investigate an example where a caregiver could fill that role in the next section.

5.3. Scaffolding Diversity: The Pool Experiment

So far, the Reuse algorithm has brought quantitative improvements in coverage, most of the time in the early phase of the exploration. We now introduce an experiment that show that Reuse can radically affect how exploration happens: namely, that can allow to explore an environment that is difficult to explore directly.

We consider the *pool task*, where two balls are present in the arena, with one out of reach. The robot must strike the first ball and make it collide with the second ball to interact with it (**Figure 15A**). The second, out-of-reach ball is the only one that is perceived through the sensory primitive. Therefore, in response to the execution of a motor command, the exploration algorithm receives the displacement of the second ball only. The exploration algorithm is therefore unaware of any interaction with the first ball.

Exploring such a task with the Explore algorithm is inefficient. The probability of interacting with the first ball during the random motor babbling phase is low (1.81%). The probability of interacting with the first ball in such a way that it collides with the second ball is very low (0.04%). Even by setting K_{boot} to 300

FIGURE 15 | The REUSE algorithm helps discover difficult-to-find areas of the sensorimotor space. In the pool experiment, a ball (in orange) is placed out of reach and the only option to interact with it is to interact with another ball (in blue), which is not tracked by the sensory primitive **(A)**. Exploring such an environment directly is difficult, because the odds of stumbling on an arm motion that leads to an interaction with the second ball are very low as demonstrated by an execution of the EXPLORE algorithm **(B)**. Using the ball task as a source **(C)**, some of those interactions are easily discovered, and the goal babbling of the pool task can start creating diversity from them **(D)**.

steps as we do for this experiment, most of the time, no interaction is witnessed after the end of random motor babbling, and goal babbling cannot function without at least one observation of a collision (**Figure 15B**).

Discovering the possibilities offered by such an environment hinges on chance. Without guidance, no informative intervention can be carried out because the environment gives neither clues about the existence of such informative interventions nor any gradient to follow toward their location: this is the bootstrapping problem, similar to the one encountered in evolutionary robotics (Mouret and Doncieux, 2009). In a context where an agent must allocate its time efficiently between different learning situations, the pooltask will most probably be quickly abandoned with the conclusion that it does not offer anything to learn.
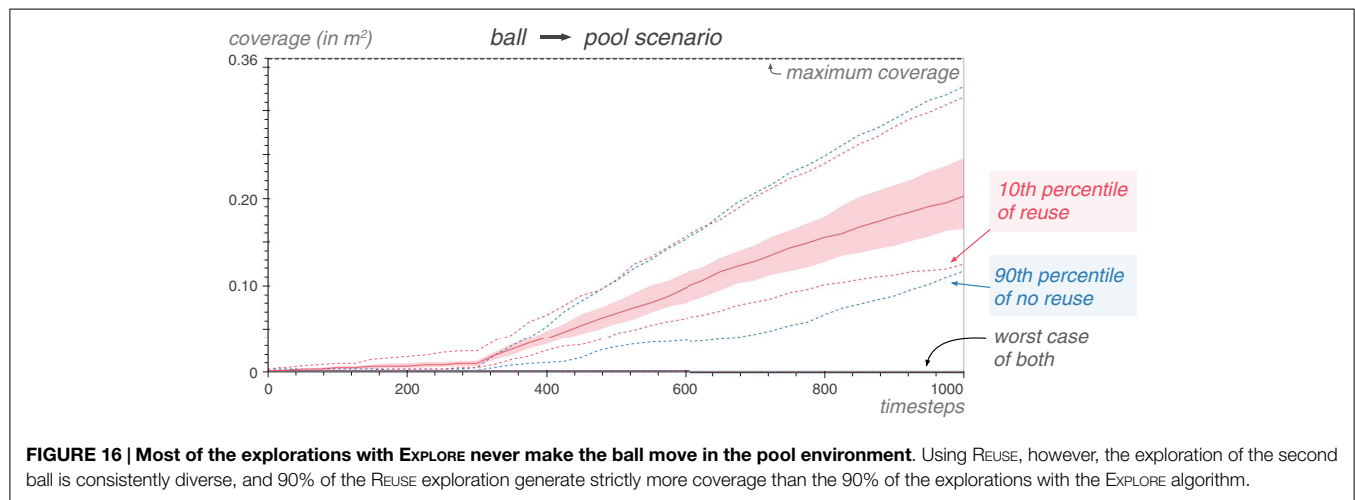
In such a context, the REUSE algorithm can provide a way to discover those interesting parts of the sensorimotor space in a reasonable amount of time. We use the ball task used in the previous sections as a source task for the pool task (**Figure 15C**). During the exploration of the ball task, the robot discovers how

to move the ball in different directions. Through REUSE, the robot replays those movements on the pool task, moving the blue ball in different directions. Some of those movements make the blue ball strike the orange ball, and thus generate novel environmental feedback. The goal babbling algorithm is then able to explore different ways the second ball can be moved (**Figure 15D**).

By looking at the coverage of the REUSE versus EXPLORE strategy over 100 repetitions of the experiment (**Figure 16**), we see that the 10th percentile of REUSE is better than the 90th percentile of the EXPLORE strategy.

This experiment showcases an important possibility offered by the REUSE algorithm: environment-driven exploration. By simply placing an agent inherently driven to explore to produce behavioral diversity, a caregiver can scaffold complex and directed behavior by manipulating the environment – here, by adding a ball – without giving any explicit goal or reward, and without the need to reprogram the robot.

The REUSE algorithm would work equally well if the source task already contained both balls, with the blue ball tracked instead

FIGURE 16 | Most of the explorations with EXPLORE never make the ball move in the pool environment. Using REUSE, however, the exploration of the second ball is consistently diverse, and 90% of the REUSE exploration generate strictly more coverage than the 90% of the explorations with the EXPLORE algorithm.

of the orange one. In that scenario, the sensory primitive would encode the attention of the robot, and moving from the source to the target task only necessitates switching the attention from the blue ball to the orange one. This is another role that a caregiver could fulfill.

5.4. Hardware Setup

In this section, we present a hybrid simulation/hardware setup that was used to validate some results of the simulation. The setup features real robots, but the interaction with the object is done in a physics engine.

5.4.1. Hybrid Interactions

The robot (**Figure 17**) has a reflective marker at the tip, which allows to accurately capture its position at 120 Hz during its movement using an OptiTrack Trio camera system. A virtual marker then *replays* the trajectory in a simulation where a virtual object has been put. As the marker is the only part of the robot tracked by the camera, it is the only part of the robotic arm that is transposed in the simulation and therefore that can collide with the object.

Contrary to the fully simulated experiment, the simulated marker does not interact with the ground where the object rest and can therefore pass through it. Moreover, the immediate reaction force on the marker can exceed 100 N without the interaction being discarded.

We chose to use a real robot and a simulated environment for the simplicity and flexibility it affords. Tracking and resetting an object a few thousands of times requires some form of mechanism, or a bigger robot, which makes the experimental setup more complicated. Additionally, the robot never experiences physical collisions, which reduces the risk of damage when babbling, given the type of robot we had. And prototyping new environments, with new objects or layouts, is cheap and unconstrained.

At the same time, using a virtual environment for an interaction task removes some of the main source of interest of the setup: a realistic, difficult to simulate interaction with a real object with kinesthetic feedback. Still, the real robot and the cameras bring real sources of motor and sensory noise that are important to check against when studying the production of diversity.

5.4.2. Cube and Ball

We reproduce the cube and ball experiments on the real setup. This time, the inverse model used is the INVERSELBFGSB-LWLR one, and the arena is 2000 mm by 2000 mm. This approximates an unbounded environment.

The results in **Figure 18** show that REUSE is effective on the hardware setup. Because the arena is more than 10 times bigger than the 600 mm by 600 mm arena, the production of coverage does not level-off as fast. In particular, the pooling around the walls seen in **Figure 10** is much less present. This explains why there is still a difference of coverage after 1000 steps between the REUSE and EXPLORE algorithm. In situations where the time allowed to explore a task is finite and much lower than would be needed to discover all the possibilities of the environment, REUSE can therefore significantly increase the amount of knowledge discovered by a robot.

5.5. Reality Gap Experiments

So far, we have shown that REUSE is effective in situations that involve switching the object (ball/cube experiment in Section 5.2), changes in the morphology of the robot (different link lengths in Section 4), or increased complexity (scaffolding experiments in Section 5.3). The purpose of using REUSE in these situations is to leverage past experiences to provide the locations of possible good mapping in the sensorimotor space.

In this section, we show that the REUSE method can be used to leverage experiences acquired in simulation on real robots, even when the simulation is not accurate.

5.5.1. The Reality Gap

Many experiments learning controllers for legged robots have reported remarkable performances for simulated robots. But far fewer have been able to transfer controllers learned in simulation onto real robots and preserve performance (Lipson et al., 2006; Palmer et al., 2009). In other words, the transfer from simulation to reality is not efficient: this is the *reality gap* problem (Jakobi et al., 1995; Jakobi, 1998). In robotics, the *reality gap* is overwhelmingly studied in the context of the optimization of controllers in simulation to be transferred on a real robot, in
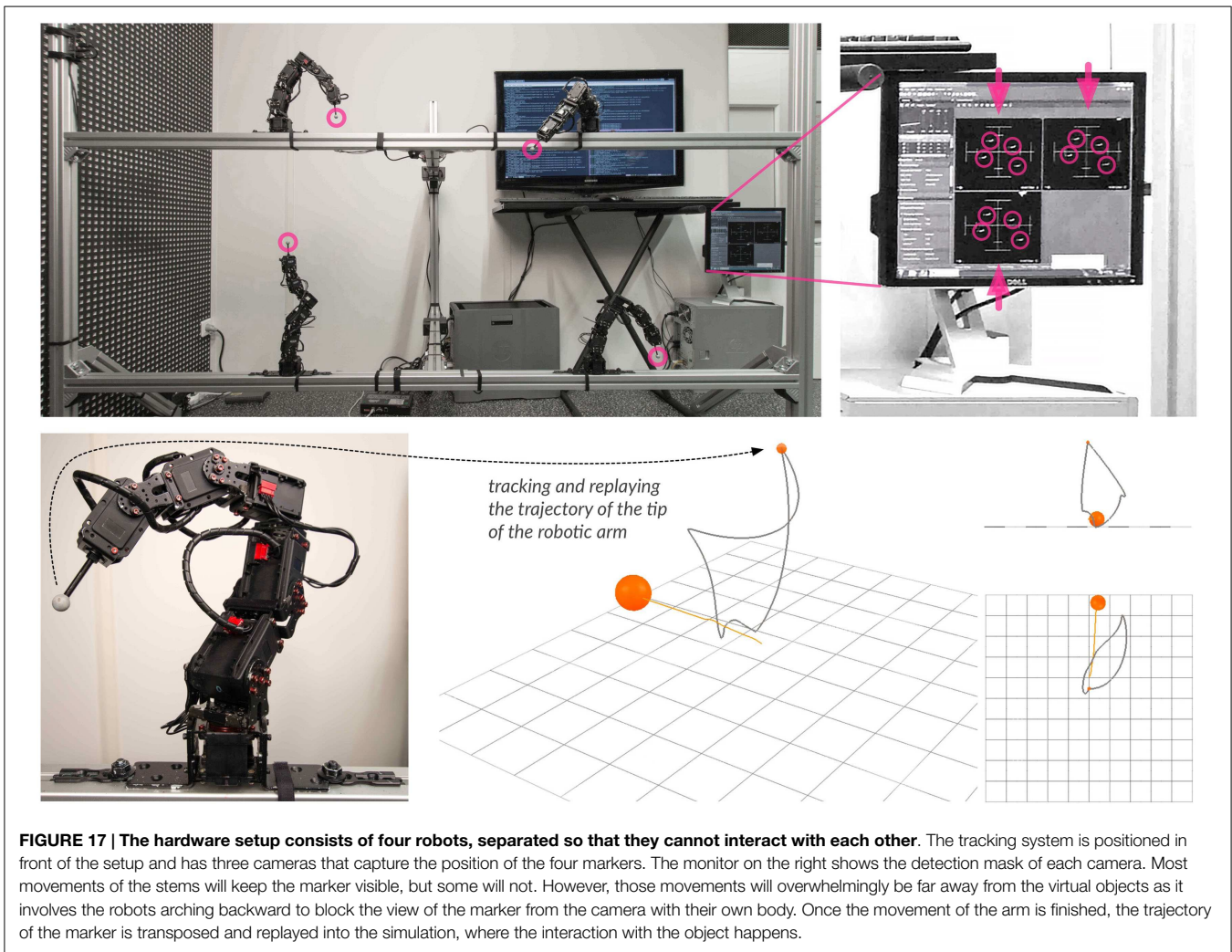
FIGURE 17 | The hardware setup consists of four robots, separated so that they cannot interact with each other. The tracking system is positioned in front of the setup and has three cameras that capture the position of the four markers. The monitor on the right shows the detection mask of each camera. Most movements of the stems will keep the marker visible, but some will not. However, those movements will overwhelmingly be far away from the virtual objects as it involves the robots arching backward to block the view of the marker from the camera with their own body. Once the movement of the arm is finished, the trajectory of the marker is transposed and replayed into the simulation, where the interaction with the object happens.

particular in the context of evolutionary robotics (Nolfi et al., 1994; Koos et al., 2013).

The most straightforward way to deal with the reality gap is to create the most accurate simulation possible. But this is fraught with problems, and leads to fragile and expensive simulations.

Some approaches improve the simulator during learning based on empirical observations (Bongard and Lipson, 2005; Bongard et al., 2006; Zagal et al., 2008; Koos et al., 2009). Other methods consider the simulator as fixed, and evaluate the mapping between the simulator and the reality. This allows to estimate the discrepancy between the two and to only perform simulated optimization in areas where the discrepancy is low (Koos et al., 2013).
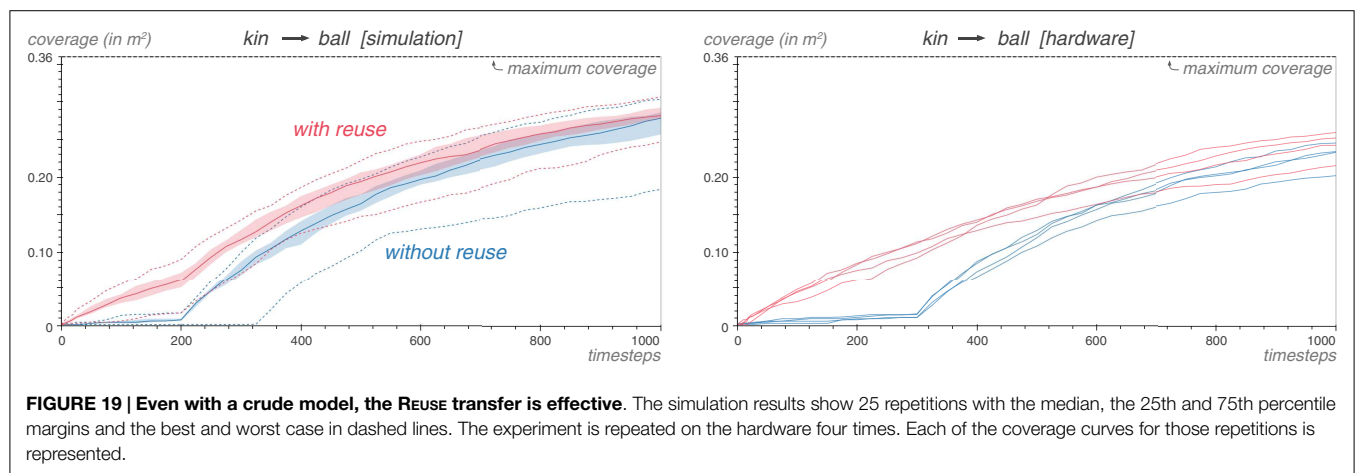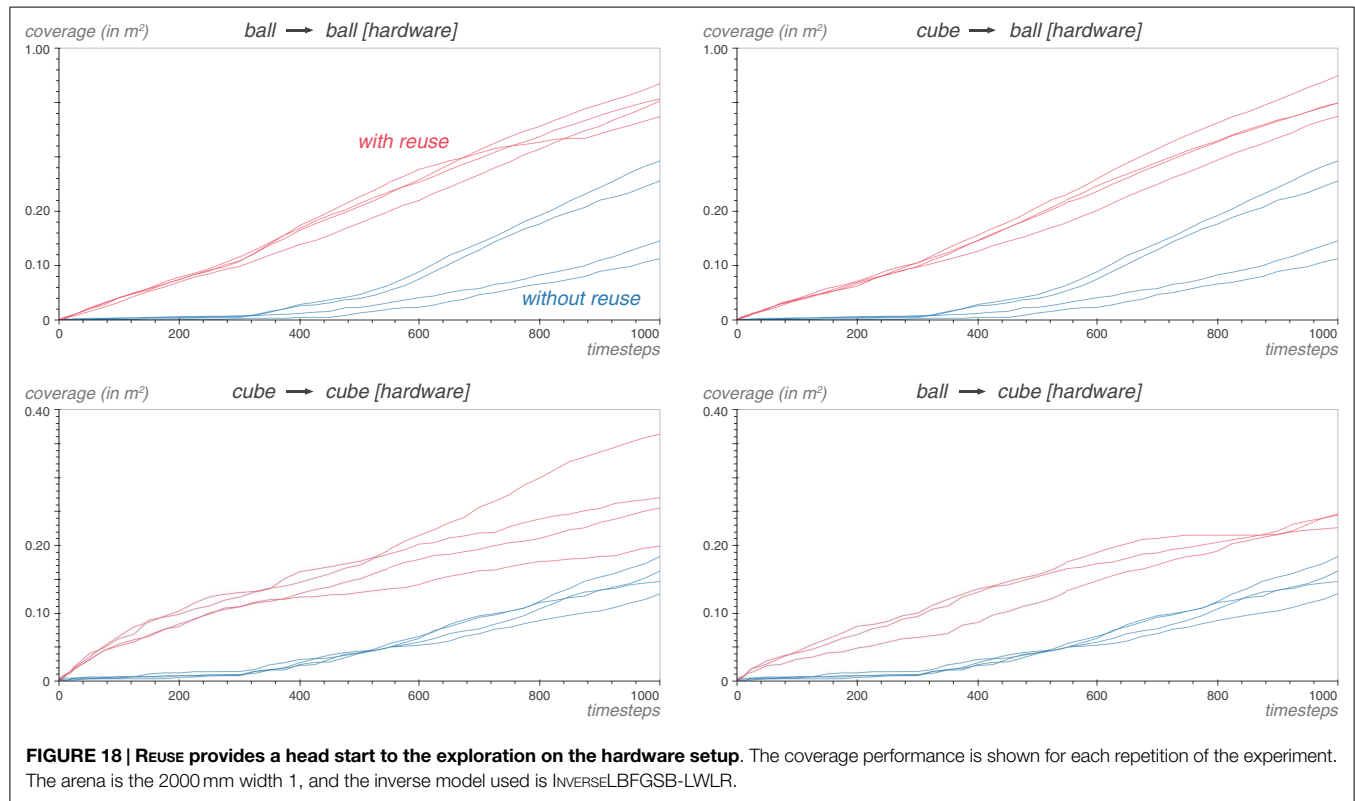
5.5.2. Crude Simulations

With REUSE, we take a different approach. Instead of spending ever-increasing efforts to create or search for a realistic simulation, we go in the opposite direction; we search for a much simpler, much cruder simulation that still affords us an exploratory advantage through REUSE. Jakobi (1997) proposes a similar method where he identifies a minimal set of features responsible for the behavior of the robot, and simulates only those. But our approach is different still: our aim is not to transfer behaviors, but it is to transfer behavioral diversity.

To test this, we create a simplified kinematic simulation of the object interaction setup of Section 5.1. Instead of using a physics engine, we compute the trajectory of the end-effector by feeding the kinematic model with the joint trajectories produced by the motor primitives. Moreover, the object is approximated to its axis-aligned bounding box. If the trajectory of the end-effector enters the bounding box, the velocity of the end-effector is averaged from its last 10 positions, and the displacement of the object is computed as a vector of the same direction as the velocity of the end-effector, and with a norm proportional to the end-effector velocity. There is no floor to interact with, the displacement of the object is computed in three dimensions, and then projected on the x and y dimensions.

Under this model, there is no difference anymore between a ball and a cube. No contact is simulated except the one between the object and the end-effector, and the collisions are computed as if they were always directed toward the center of mass of the object.

The kinematic simulation is run for 1000 timesteps using the EXPLORE exploration strategy ($K_{boot} = 300$, $d = 0.05$). The exploration is then transferred to the V-REP simulation of the ball task of Section 5.1. The exploration on the full simulation uses the REUSE algorithm and is parametrized with $K_{boot} = 300$, $d = 0.05$, and $p_{reuse} = 50\%$. The results are available **Figure 19**.

FIGURE 18 | Reuse provides a head start to the exploration on the hardware setup. The coverage performance is shown for each repetition of the experiment. The arena is the 2000 mm width 1, and the inverse model used is InverseLBFGSB-LWLR.

FIGURE 19 | Even with a crude model, the Reuse transfer is effective. The simulation results show 25 repetitions with the median, the 25th and 75th percentile margins and the best and worst case in dashed lines. The experiment is repeated on the hardware four times. Each of the coverage curves for those repetitions is represented.

Even with a crude simulation devoid of most physical features, the Reuse strategy is able to take significant advantage of the generated data.

5.5.3. A Cruder Simulation

We simplify the previous simulation. Instead of computing the displacement of the object, the sensory response is only conditioned by the end-effector entering the bounding box. If that happens, a *random* value between 0 and 1 is returned. Else, a random value between −1 and 0 is returned. The sensory signal has only one dimension. This experiment also affords us with another example of Reuse being compatible with a change in sensory modalities.

Learning with such a poor sensory feedback is more difficult. The simulation has essentially become an indicator for a possible collision. Yet, Reuse still provides an improvement (**Figure 20**). As should be expected, the improvement is less than when the simulation is more informative.

A weakness of our reality gap experiments is that even a simple forward kinematic model usually displays good performance on a rigid body robotic arm. Although we removed many aspects of the physical simulation, we retained the essential part. The discrepancy then between a collision detected in simulation and one produced in reality is low. This easily explains the results obtained. And while we claimed not to assume that the simulation needs to be physically accurate, it actually is, but qualitatively.

FIGURE 20 | Even with a cruder model, the REUSE transfer is still effective. 25 repetitions.
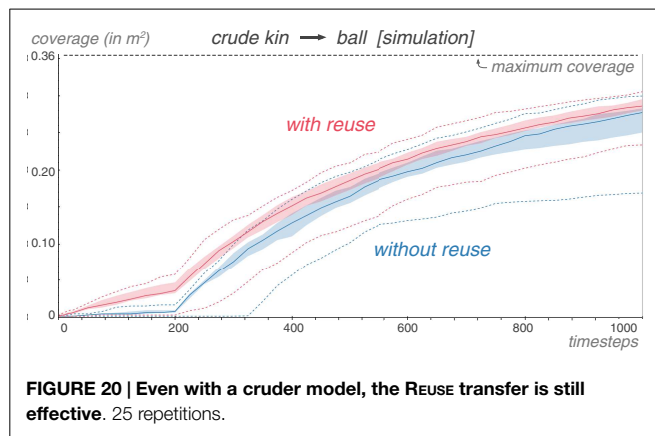
The way the object displacement is computed in the first crude simulation can also be criticized. Although it seems that, by not taking into account any geometry of the object, or not considering the floor we have lost much information, the direction of the displacement is directly correlated to the direction of the end-effector when a collision happens. This sensory feedback is probably richer in information that the final position of the object in the physical simulation. It is also a signal that is easier to learn. The first crude simulation could be considered as a scaffolding that offers knowledge of a pivotal aspect of the interaction – the direction and velocity of the colliding tip of the arm just before the collision – that was hidden so far.

Of course, these criticisms can also be considered positively: yes, the crude models are qualitatively accurate with regards to the presence of a collision, and REUSE is able to take advantage of a merely qualitative, rather than numerical, accuracy.

In a self-sufficient perspective, the crude simulations could be considered as cognitive models. Their simplicity and relaxed qualitative nature makes their acquisition by a self-sufficient robot more reasonable than realistic simulations. Instead of reproducing reality, these cognitive simulations can do away with much of the realism while retaining power to direct and inform behavior. They pose as artifices of cognition that would allow robots, in some situations, to reason about the world without having to predict or simulate it accurately.

6. RELATED WORKS

Goal-directed exploration (Oudeyer and Kaplan, 2007; Baranes and Oudeyer, 2010; Jamone et al., 2011; Rolf et al., 2011; Baranes and Oudeyer, 2013), as well as related methods such as MAP-Elite (Cully et al., 2015), has been shown to be effective at creating behavioral diversity in large sensorimotor spaces. However, these methods only consider a single task. The REUSE algorithm proposes to transfer the behavioral diversity from one task to another. It, therefore, works particularly well when combined with these strategies as we have demonstrated in this paper.

The REUSE method is an instance of a transfer algorithm. Machine learning algorithms improve their prediction or control capabilities from data. Transfer learning algorithms (Thrun and Pratt, 1998; Taylor and Stone, 2009; Pan and Yang, 2010) aim at improving their prediction or control capabilities on a problem

either from another problem's existing data or more directly from the other problem's learned prediction or control capabilities. In other words, transfer learning expands the scope of the data that can be used on a specific problem.

Therefore, transfer learning is typically used when not enough data is available to obtain the desired performance. Creating a zebra classifier can be difficult if only a few labeled pictures are available. While a horse classifier does not address the exact same problem, enough commonalities exist between the two for useful information to be extracted from the horse classifier and used in the zebra one.

Transfer learning algorithms have been historically developed for tasks where unlabeled data is plentiful, but labeling is expensive; robots face a similar labeling problem. Every motor action a robot undertakes is costly in time and energy. Therefore, while the motor action possibilities are numerous, only a small fraction of them can be executed to observe the environmental response they produce – i.e., labeled – during the time allotted to learn a problem. Transfer learning in robots allows to make use of the observations acquired outside of the current problem.

Many different methods have been developed on how, what, and when to transfer data from one task to another, and the interested reader can consult Thrun and Pratt (1998), Taylor and Stone (2009), Pan and Yang (2010), and Lazaric (2012) for reviews.

In evolutionary robotics, Velez and Clune (2014) shows that controllers evolved in a first maze through Novelty Search (Lehman and Stanley, 2011), i.e., with the incentive to behave differently from the rest of the population, provide a head start on the exploration of a second maze. In comparison with REUSE, the transferred controllers are valued not because they solve different tasks, or explore the maze differently: they are issued from independent runs, and all solve the same task, going to the same predefined goal. Rather, they are valued because they can adapt faster to the new task than random controllers, having acquired exploration abilities in the first maze. In the Intelligent Trial and Error algorithm (IT&E) (Cully et al., 2015), a performance map is generated by MAP-Elite on a task and then reused to find a fast adaptation to a different task. However, the performance mapping is used to guide the search, which is focused on a specific objective.

In reinforcement learning, an interesting method comes from Sherstov and Stone (2005) that creates a set of tasks from a source task, and prune the action space from any action that is not optimal in at least one task of the set. The diversity of the set of tasks creates a filter that is used to reduce diversity in the set of actions.

In the context of Markov-Decision Processes (MDP), *policy reuse* (Fernández and Veloso, 2006) builds a library of policy. Each policy corresponds to a specific reward function over the MDP. Each time a new reward function needs to be learned, the most similar policy in the policy library is reused probabilistically with a ε-greedy strategy. The policy reuse algorithm focuses on learning how to solve a single reward function at a time, over discrete or discretized domains. Like IT&E, it uses the reward function to decide which policy to reuse.

We first exposed the REUSE method in Benureau and Oudeyer (2013). The TRANSFER was driven by intrinsic motivation then. It was changed to a diversity-driven method in

Benureau et al. (2014) but the empirical results presented then were limited to the content of Sections 5.2.3 and 5.4.2. This paper provides a comprehensive empirical study and investigates many different situations: changes in morphology, sensory modalities, and the exploitation of a random motor babbling source. The effectiveness of Reuse is explained by showing how the diversity of two different tasks can be highly correlated, and we investigates the details of a situation where Reuse, at his worst, is less efficient than the worst-case without it. The paper also makes two major new contributions: the application of Reuse to scaffolding behavior (Section 5.3) and exploiting simulation results on real robots (Section 5.5).

7. DISCUSSION

7.1. Synthesis

Sensorimotor spaces present difficulties that preclude an isolated approach. They typically feature a large motor space that cannot be explored exhaustively. Rather, often, only a few small regions of it are actually interesting to explore for any practical purpose. The difficulty, of course, lies in *discovering* those regions. However smart an exploration algorithm is, when the environment does not provide clues or gradient toward those regions, finding them relies on chance.

The Reuse method proposes a way to discover these small regions of interesting behavior by relying on past experience. In an autonomous context where neither experts nor peers are present, and in a developmental context where robots are supposed to accumulate experience about the world over large periods of time, relying on past experience seems trivially self-evident. It can prove, however, challenging. A strength of the Reuse method is that it makes easy to use past experiences that would otherwise be considered incompatible with the current situation. We have provided examples of the Reuse method adapting to changes in objects (cube and ball, Section 5.2), in morphology (length of arm links in Section 4), in task dissimilarity (change in the ball position, Section 5.2.3), in sensory modalities (coordinate systems in Section 4 and crude simulation in Section 5.5.3), in complexity (pool experiment in Section 5.3), and in execution context (from simulation to reality in Section 5.5).

Yet, the Reuse method remains, at its heart, remarkably simple: create a collection of actions having generated a diversity of effects in a previous task, and optimistically reexecute them in the new one. As a consequence, the method is algorithmically cheap. The only constraint is that actions from the source task must be reexecutable in the target task.

While the Reuse method makes many past experiences suddenly compatible with the current situation, it does not mean that they are relevant or beneficial. The planar arms experiments (Section 4) provided us with evidence that complex interactions between the tasks, the inverse model and the Reuse method may worsen the exploration in some cases rather than help it. And much of the success of the Reuse method lies in the similarity between the tasks. When the two tasks are too dissimilar, the Reuse method needs to degrade gracefully, and this is what the experiment with the displaced task demonstrated (Section 5.2.3).

The Reuse method does not merely improve or accelerate the exploration of sensorimotor spaces. As the pool experiment illustrates (Section 5.3), it can scaffold the exploration of difficult environments. It allows a caregiver to guide the exploration of an autonomous agent, leading it to acquire specific and sophisticated behaviors, without specifying an explicit goal or reward, by either directing the attention or manipulating the environment.

The Reuse method also seems naturally suited to propose solutions to a difficult problem: exploiting simulation results on real robots. The Reuse method does this by side-stepping the difficulty of preserving performance. Instead, it focuses on preserving behavioral diversity, providing good starting points in the sensorimotor space of the real robot (Section 5.5). Moreover, our experiment with crude simulations suggests that cheap cognitive models can efficiently serve an efficient source of behavioral diversity, informing the exploration in the real world and in full-featured simulations.

7.2. Limitations and Perspectives

The works presented here suffers many limitations. The Reuse algorithm is only analyzed with regards to the behavioral diversity it creates through the τ-coverage metrics and not on the quality of either the predictive or control models that can be derived from the observations it generates. This should be investigated in future works, especially in context where robots must apply their skills and knowledge to reach specific goals.

This leads us to the issue of chaotic environments, evoked in Section 3. In the full simulation of the interaction task (Section 5.1), we monitored the reactive forces to mitigate the chaotic behavior of the physics engine. More generally, chaotic areas of the sensorimotor spaces generate behavioral diversity that is difficult to exploit for practical purposes. To make Reuse more robust to these aspects, the chaotic and stochastic characteristics of the reused motor command should be explicitly evaluated.

Another blind spot of the Reuse algorithm is motor command diversity. When the effect diversity of the source task is low, motor commands producing similar effects are reused. In such a case, choosing motor commands according to how different they are from one another would increase the diversity of the set of transferred commands. This would make Reuse robust to scenarios where, for instance, all motor commands produce the same effect in the source task.

Another venue of improvement would be to make Reuse active, i.e., aware of its own effect of the exploration. By having a feedback on its performance, the algorithm could dynamically decide to modify the value of p_{reuse} or of the length of the bootstrapping phase (Section 5.2.2). This could also lead to Reuse identifying which parts of the source task observations produce the most diversity on the target tasks, and to preferentially to Reuse motor commands from those areas: this could have been exploited in the pool experiment in particular (Section 5.3).

In this paper, we have only considered one source task. But Reuse should be expanded to multiple tasks scenarios, since autonomous developmental robots are not expected to have only one explored task in their past experience.

The experimental setups presented in this paper do not yet allow to generalize to many robotic contexts. The algorithm

should be tested in more diverse environments, with different sensory primitives, and in particular fully autonomous, real-world ones.

AUTHOR CONTRIBUTIONS

Ideas and design: FB (75%) and P-YO (25%), experiments: FB (100%), code: FB (100%), writing: FB (75%), and P-YO (25%).

ACKNOWLEDGMENTS

Thanks to Paul Fudal, whose engineering expertise made the experiments on object interaction possible. Thanks to Thomas Cederborg for his helpful comments during the redaction of the manuscript.

REFERENCES

Atkeson, C. G., Moore, A. W., and Schaal, S. (1997a). "Locally weighted learning," in *Lazy Learning*, ed. D. W. Aha (Dordrecht: Springer), 11–73.

Atkeson, C. G., Moore, A. W., and Schaal, S. (1997b). "Locally weighted learning for control," in *Lazy Learning*, ed. D. W. Aha (Dordrecht: Springer), 75–113.

Baldassarre, G., and Mirolli, M. (eds) (2013). *Intrinsically Motivated Learning in Natural and Artificial Systems*. Berlin, Heidelberg: Springer Science + Business Media.

Baranes, A., and Oudeyer, P.-Y. (2013). Active learning of inverse models with intrinsically motivated goal exploration in robots. *Rob. Auton. Syst.* 61, 49–73. doi:10.1016/j.robot.2012.05.008

Baranes, A., and Oudeyer, P.-Y. (2010). "Intrinsically motivated goal exploration for active motor learning in robots: a case study," in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (Taipei: IEEE), 1766–1773.

Benureau, F., Fudal, P., and Oudeyer, P.-Y. (2014). "Reusing motor commands to learn object interaction," in *Proceedings of ICDL-Epirob 2014*, IEEE, Genoa.

Benureau, F., and Oudeyer, P.-Y. (2013). "Autonomous reuse of motor exploration trajectories," in *Proceedings of ICDL-Epirob 2013*, 1–8.

Benureau, F., and Oudeyer, P.-Y. (2015). "Diversity-driven selection of exploration strategies in multi-armed bandits," in *ICDL-EPIROB 2015*, Providence, RI.

Benureau, F., and Oudeyer, P.-Y. (2016). Code and data for "Behavioral Diversity Generation in Autonomous Exploration Through Reuse of Past Experience". doi:10.6084/m9.figshare.2816284.v1

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., and Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cogn. Psychol.* 64, 215–234. doi:10.1016/j.cogpsych.2011.12.002

Bongard, J., and Lipson, H. (2005). Nonlinear system identification using coevolution of models and tests. *IEEE Trans. Evol. Comput.* 9, 361–384. doi:10.1109/tevc.2005.850293

Bongard, J., Zykov, V., and Lipson, H. (2006). Resilient machines through continuous self-modeling. *Science* 314, 1118–1121. doi:10.1126/science.1133687

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* 16, 1190–1208. doi:10.1137/0916069

Cazals, F., Kanhere, H., and Loriot, S. (2011). Computing the volume of a union of balls. *ACM Trans. Math. Softw.* 38, 1–20. doi:10.1145/2049662.2049665

Cleveland, W. S., and Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* 83, 596–610. doi:10.1080/01621459.1988.10478639

Cook, C., Goodman, N. D., and Schulz, L. E. (2011). Where science starts: spontaneous experiments in preschoolers' exploratory play. *Cognition* 120, 341–349. doi:10.1016/j.cognition.2011.03.003

Cully, A., Clune, J., Tarapore, D., and Mouret, J.-B. (2015). Robots that can adapt like animals. *Nature* 521, 503–507. doi:10.1038/nature14422

Fernández, F., and Veloso, M. (2006). "Probabilistic policy reuse in a reinforcement learning agent," in *Proceedings of AAMAS 2006*, AAMAS '06, ACM, 720–727.

Gopnik, A. (1997). *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.

FUNDING

SUPPLEMENTARY MATERIAL

The source code and data for reproducing the experiments and producing all graphs, as well as the provenance data of the experiments and instructions to recreate the computational environments is published (Benureau and Oudeyer, 2016) and is made available at: https://dx.doi.org/10.6084/m9.figshare.2816284

Gopnik, A. (2012). Scientific thinking in young children: theoretical advances, empirical research, and policy implications. *Science* 337, 1623–1627. doi:10.1126/science.1223416

Gopnik, A., Sobel, D. M., Schulz, L. E., and Glymour, C. (2001). Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Dev. Psychol.* 37, 620–629. doi:10.1037/0012-1649.37.5.620

Gweon, H., and Schulz, L. (2008). "Stretching to learn: ambiguous evidence and variability in preschooler's exploratory play," in *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (Austin, TX: Cognitive Science Society), 1552–1556.

Hoffmann, M., and Pfeifer, R. (2012). *The Implications of Embodiment for Behavior and Cognition: Animal and Robotic Case Studies*. ArXiv e-prints.

Ijspeert, A. J., Nakanishi, J., Hoffmann, H., Pastor, P., and Schaal, S. (2013). Dynamical movement primitives: learning attractor models for motor behaviors. *Neural Comput.* 25, 328–373. doi:10.1162/neco_a_00393

Ijspeert, A. J., Nakanishi, J., and Schaal, S. (2002). "Movement imitation with nonlinear dynamical systems in humanoid robots," in *Proceedings of the 2002 IEEE International Conference on Robotics and Automation (ICRA)*, (Washington, DC: IEEE), 1398–1403.

Jakobi, N. (1997). Evolutionary robotics and the radical envelope-of-noise hypothesis. *Adapt. Behav.* 6, 325–368. doi:10.1177/105971239700600205

Jakobi, N. (1998). "Running across the reality gap: octopod locomotion evolved in a minimal simulation," in *Evolutionary Robotics: First European Workshop, EvoRobot98 Paris, France, April 16–17, 1998 Proceedings*, eds P. Husbands and J.-A. Meyer (Springer Science + Business Media), 39–58.

Jakobi, N., Husbands, P., and Harvey, I. (1995). "Noise and the reality gap: the use of simulation in evolutionary robotics," in *Advances in Artificial Life: Third European Conference on Artificial Life Granada, Spain, June 4–6, 1995 Proceedings*, eds F. Morán, A. Moreno, J. J. Merelo, and P. Chacón (Berlin, Heidelberg: Springer Science + Business Media), 704–720.

Jamone, L., Natale, L., Hashimoto, K., Sandini, G., and Takanishi, A. (2011). "Learning task space control through goal directed exploration," in *IEEE International Conference on Robotics and Biomimetics (ROBIO)* (Karon Beach: IEEE), 702–708.

Koos, S., Mouret, J.-B., and Doncieux, S. (2009). "Automatic system identification based on coevolution of models and tests," in *Proceedings of the 2009 IEEE Congress on Evolutionary Computation, (CEC)* (Trondheim), 560–567.

Koos, S., Mouret, J.-B., and Doncieux, S. (2013). The transferability approach: crossing the reality gap in evolutionary robotics. *IEEE Trans. Evol. Comput.* 17, 122–145. doi:10.1109/tevc.2012.2185849

Kulvicius, T., Ning, K., Tamosiunaite, M., and Worgötter, F. (2012). Joining movement sequences: modified dynamic movement primitives for robotics applications exemplified on handwriting. *IEEE Trans. Robot.* 28, 145–157. doi:10.1109/TRO.2011.2163863

Lazaric, A. (2012). "Transfer in reinforcement learning: a framework and a survey," in *Reinforcement Learning: State-of-the-Art*, eds M. Wiering, and M. Otterlo (Berlin, Heidelberg: Springer Science + Business Media), 143–173.

Lehman, J., and Stanley, K. O. (2011). Abandoning objectives: evolution through the search for novelty alone. *Evol. Comput.* 19, 189–223. doi:10.1162/evco_a_00025

Lipson, H., Bongard, J., Zykov, V., and Malone, E. (2006). "Evolutionary robotics for legged machines: from simulation to physical reality," in *Proceedings of the 9th International Conference on Intelligent Autonomous Systems*, 11–18.

Loeb, G. E. (2012). Optimal isn't good enough. *Biol. Cybern.* 106, 757–765. doi:10.1007/s00422-012-0514-6

Lorenz, K. (1996). "Innate bases of learning," in *Learning as Self-Organization*, eds K. H. Pribram and J. S. King (Mahwah, NJ: L. Erlbaum Associates), 1–56.

Morales, J. L., and Nocedal, J. (2011). Remark on "algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization." *ACM Trans. Math. Softw.* 38, 1–4. doi:10.1145/2049662.2049669

Moulin-Frier, C., Nguyen, S. M., and Oudeyer, P.-Y. (2014). Self-organization of early vocal development in infants and machines: the role of intrinsic motivation. *Front. Psychol.* 4:1006. doi:10.3389/fpsyg.2013.01006

Moulin-Frier, C., and Oudeyer, P.-Y. (2013). "Exploration strategies in developmental robotics: a unified probabilistic framework," in *IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)*. Osaka: IEEE.

Mouret, J.-B., and Doncieux, S. (2009). "Overcoming the bootstrap problem in evolutionary robotics using behavioral diversity," in *IEEE Congress on Evolutionary Computation* (Trondheim: IEEE), 1161–1168.

Munzer, T., Stulp, F., and Sigaud, O. (2014). "Non-linear regression algorithms for motor skill acquisition: a comparison," in *Proceedings JFPDA* (Liège), 1–16.

Nolfi, S., Floreano, D., Miglino, O., and Mondada, F. (1994). "How to evolve autonomous robots: different approaches in evolutionary robotics," in *Artificial life IV: Proceedings of the 4th International Workshop on Artificial Life* (MA: MIT Press), 190–197.

Oudeyer, P.-Y., and Kaplan, F. (2007). What is intrinsic motivation? a typology of computational approaches. *Front. Neurorobot.* 1:6. doi:10.3389/neuro.12.006.2007

Palmer, M., Miller, D., and Blackwell, T. (2009). "An evolved neural controller for bipedal walking: Transitioning from simulator to hardware," in *Proceedings of IROS 2009 Workshop on Exploring New Horizons in Evolutionary Design of Robots*.

Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi:10.1109/TKDE.2009.191

Penrose, R., and Todd, J. A. (1955). A generalized inverse for matrices. *Math. Proc. Cambridge Philos. Soc.* 51, 406. doi:10.1017/s0305004100030401

Rolf, M., Steil, J., and Gienger, M. (2011). "Online goal babbling for rapid bootstrapping of inverse models in high dimensions," in *Proceedings of ICDL 2011*, Vol. 2, (Frankfurt am Main), 1–8.

Schulz, L. E., and Bonawitz, E. B. (2007). Serious fun: preschoolers engage in more exploratory play when evidence is confounded. *Dev. Psychol.* 43, 1045–1050. doi:10.1037/0012-1649.43.4.1045

Schulz, L. E., Gopnik, A., and Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Dev. Sci.* 10, 322–332. doi:10.1111/j.1467-7687.2007.00587.x

Sherstov, A. A., and Stone, P. (2005). "Improving action selection in mdp's via knowledge transfer," in *Proceedings of the 20th National Conference on Artificial Intelligence, AAAI'05*, Vol. 2, Pittsburgh: AAAI Press.

Stulp, F. (2014). *DmpBbo – A C++ Library for Black-Box Optimization of Dynamical Movement Primitives*.

Taylor, M. E., and Stone, P. (2009). Transfer learning for reinforcement learning domains: a survey. *J. Mach. Learn. Res.* 10, 1633–1685.

Thrun, S., and Pratt, L. (eds) (1998). *Learning to Learn*. (Boston, MA: Springer Science + Business Media), 3–17. doi:10.1007/978-1-4615-5529-2

Till, M. S., and Ullmann, G. M. (2009). McVol – a program for calculating protein volumes and identifying cavities by a Monte Carlo algorithm. *J. Mol. Model.* 16, 419–429. doi:10.1007/s00894-009-0541-y

Velez, R., and Clune, J. (2014). "Novelty search creates robots with general skills for exploration," in *Proceedings of the 2014 Conference on Genetic and Evolutionary Computation – GECCO '14* (Vancouver, BC: ACM), 737–744.

Zagal, J. C., Ruiz-del Solar, J., and Palacios, A. G. (2008). "Fitness based identification of a robot structure," in *Artificial Life XI: Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems*, eds S. Bullock, J. Noble, R. Watson, and M. A. Bedau (Cambridge, MA: MIT Press), 733–740.

Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J. (1997). Algorithm 778: L-BFGS-B: fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.* 23, 550–560. doi:10.1145/279232.279236

Online Body Schema Adaptation Based on Internal Mental Simulation and Multisensory Feedback

Pedro Vicente*, Lorenzo Jamone and Alexandre Bernardino

Institute for Systems and Robotics (ISR/IST), LARSyS, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

In this paper, we describe a novel approach to obtain automatic adaptation of the robot body schema and to improve the robot perceptual and motor skills based on this body knowledge. Predictions obtained through a mental simulation of the body are combined with the real sensory feedback to achieve two objectives simultaneously: body schema adaptation and markerless 6D hand pose estimation. The body schema consists of a computer graphics simulation of the robot, which includes the arm and head kinematics (adapted online during the movements) and an appearance model of the hand shape and texture. The mental simulation process generates predictions on how the hand will appear in the robot camera images, based on the body schema and the proprioceptive information (i.e., motor encoders). These predictions are compared to the actual images using sequential Monte Carlo techniques to feed a particle-based Bayesian estimation method to estimate the parameters of the body schema. The updated body schema will improve the estimates of the 6D hand pose, which is then used in a closed-loop control scheme (i.e., visual servoing), enabling precise reaching. We report experiments with the iCub humanoid robot that support the validity of our approach. A number of simulations with precise ground-truth were performed to evaluate the estimation capabilities of the proposed framework. Then, we show how the use of high-performance GPU programing and an edge-based algorithm for visual perception allow for real-time implementation in real-world scenarios.

Keywords: humanoid robot, internal learning model, visual control, simulation, body schema

1. INTRODUCTION

Humans develop body awareness through an incremental learning process that starts in early infancy (von Hofsten, 2004), and probably even prenatally (Joseph, 2000). Such awareness is supported by a neural representation of the body that is constantly updated with multimodal sensorimotor information acquired during motor experience and that can be used to infer the limbs' position in space and guide motor behaviors: a body schema (Berlucchi and Agliotti, 1997).

In particular, during the first months of life, infants spend a considerable amount of time observing their own hands while moving (Rochat, 1998). Specific experiments in which babies were laying supine in the dark, with the head turned to one side, show voluntary arm control to bring the hand within the cone of light emitted by a narrow beam, so to make it visible (Van der Meer, 1997). These early behaviors might support an initial visual-proprioceptive calibration of the eye-hand system, which is required to perform reaching movements effectively later on.

Indeed, while until 4 months reaching movements seem not to exploit any visual feedback, as trajectory correction is absent (Bushnell, 1985; von Hofsten, 1991), from 5 months vision is used to correct the hand pose during the movement (Mathew and Cook, 1990), with performance that improves incrementally (Ashmead et al., 1993). However, after 9 months, this visual guidance almost disappears, as children become able to plan a proper hand trajectory at the movement onset (Lockman et al., 1984). According to Bushnell (1985), this decline of visually guided reaching is fundamental for the further cognitive development of the infant, as it frees a big portion of visual attention that can be, thus, devoted to perceive and learn other aspects of the experienced situations.

Interestingly, these observations suggest that an internal model might have been learned through sensorimotor experience during the first months, and later exploited for improved motion control. Indeed, general theories of human motor learning and control claim that forward and inverse internal models of the limbs are learned and maintained updated in the cerebellum (Wolpert et al., 1998). While inverse models are used to compute the muscle activations required to perform a desired movement, forward models can be used to simulate motor behaviors and to predict the sensory outcomes of specific movements (Miall and Wolpert, 1996). These predictions are exploited in different ways: for example, they are combined with the actual sensory feedback through Bayesian integration to improve the estimation of the current state of the system (Körding and Wolpert, 2004).

Moreover, according to Sober and Sabes (2005), humans use visual and proprioceptive signals to estimate the position of the arm during the planning of reaching movements. The combination of these two feedback sources is dependent not only on the task but also on the content of the visual information, suggesting a strategy to minimize the predictive error. The brain chooses the best combination to reduce the influence of the noise present in the feedback signals.

Clearly, endowing artificial agents with similar capabilities is a major challenge for cognitive robotics, and it paves the way for the next generation of autonomous humanoid robots that will have to operate alongside humans in unstructured environments.

Fundamental tasks like grasping objects while avoiding obstacles and self-collisions require an accurate representation of the body schema. For example, think about an apparently simple task like taking a coffee mug and give it to a human without spilling out the content: a precise estimation and control of the end-effector pose are of paramount importance both to first approach the mug and grasp it and to continuously control its pose during the movement.

The robots employed in very structured environment (e.g., industrial robots) might not need to use vision for similar tasks (i.e., the objects are already in known positions), or might use images coming from cameras that are fixed in the environment. In these cases, a calibration of the system performed during occasional maintenance operations is typically enough to guarantee the repeatability of the movements. Instead, humanoid robots are complex systems with many moving parts, including the cameras providing the visual inputs, which are typically located in a moving head: for this kind of systems, a continuous online recalibration is needed to assure the accuracy of visually guided movements. Moreover, robot vision in unstructured environments is more challenging because of the unpredictable nature of the image background: strategies to cope with this kind of visual feedback are, therefore, required.
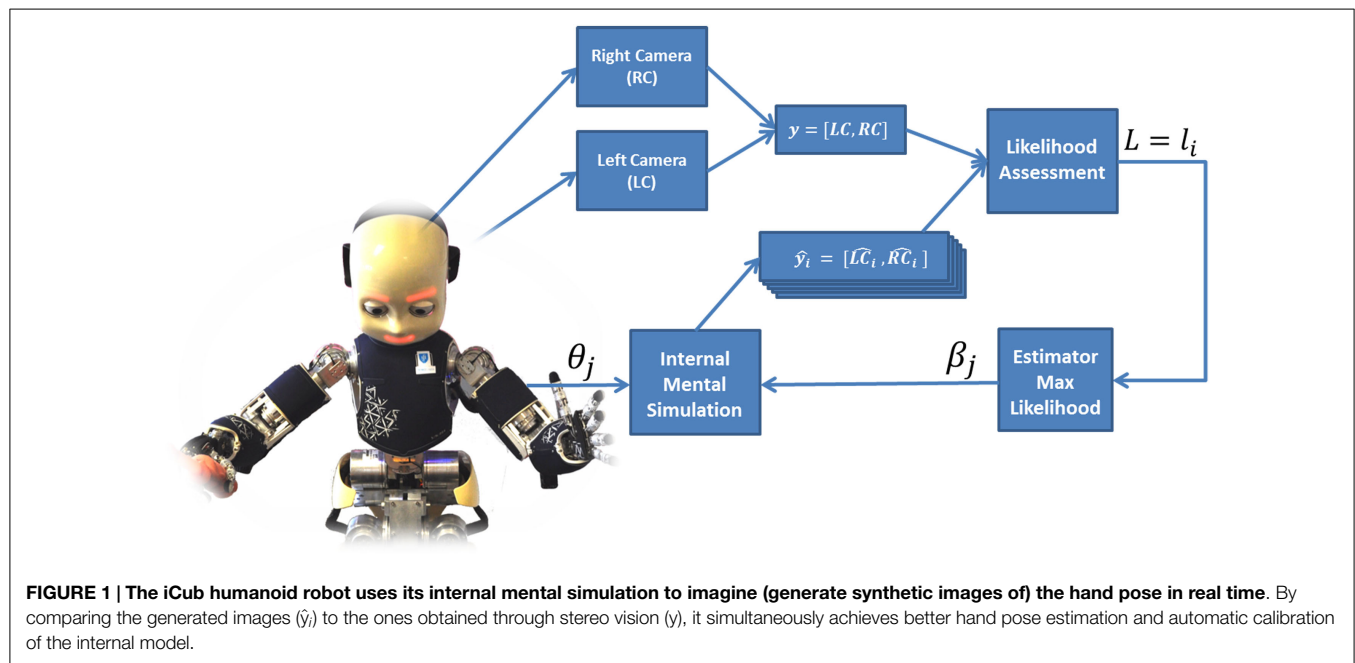
Our objective in this paper is to perform continuous online adaptation of an analytical internal model of the robot (i.e., the robot body schema) using multimodal sensory information (i.e., vision and proprioception), and to exploit the updated model to facilitate the estimation of the 6D pose of the robot end-effector (i.e., the hand palm).

Some works have been proposed in the literature to address the body schema adaptation problem (See Related Work). However, most of them rely on artificial markers to visually identify the robot end-effector, and use local optimization method for the model adaptation. Our method goes further by using natural visual cues (marker-free solution) and a global estimation approach based on sequential Monte Carlo methods for the model adaptation.

We apply our method to the iCub humanoid robot (Metta et al., 2010), depicted in **Figure 1**. Instead of learning an internal model from scratch, we exploit a computer graphics (CG) model of the robot that includes the CAD kinematics [provided within the YARP/iCub software framework as described in Pattacini (2011)] and an appearance model of the hand shape and texture. This model is adapted in real-time during reaching movements using data from the motor encoders and from the cameras located in the robot eyes. The model adaptation consists in estimating a set of joint offsets to be added to the CAD kinematics to better describe the real robot. The model is a forward model, and it can be directly used to make forward predictions. In our approach, we use the model and the encoders measurements to make predictions (i.e., visual hypotheses) about how the hand should appear in the robot cameras; the predictions are combined with the actual visual information using Bayesian techniques (i.e., sequential Monte Carlo) to estimate the 6D pose of the end-effector (3D position and 3D orientation) and to calibrate the model kinematics. Moreover, the model can be used to make inverse predictions, which are required for movement control. In particular, we implement both feedforward control (open-loop), based on inverse kinematics computation, and feedback control (closed-loop), using the pseudo-inverse of the model Jacobian and the estimated pose of the end-effector as the feedback signal.

We report experiments both in simulation, with the iCub dynamic simulator (Tikhanoff et al., 2008), and with the real iCub humanoid robot. The real-time implementation on the real robot is made possible by two techniques: GPU programing, to achieve faster computation, and an edge-based metric to compare the visual hypothesis with the actual visual perception. This is essential to improve robustness in real-world scenes, with a natural non-structured background.

Our solution draws inspiration from human development and learning, as: (i) the internal model is updated online based on the visual feedback of the hand, as infants seem to do between 4 and 8 months of age and (ii) the estimation of the pose of the end-effector results from the Bayesian integration of the sensory (visual) feedback and the predictions made by the internal model,

FIGURE 1 | The iCub humanoid robot uses its internal mental simulation to imagine (generate synthetic images of) the hand pose in real time. By comparing the generated images ($\hat{y}_i$) to the ones obtained through stereo vision (y), it simultaneously achieves better hand pose estimation and automatic calibration of the internal model.

a strategy that seems to characterize human perception as well, as described in Körding and Wolpert (2004).

The rest of the paper is organized as follows. In Section "Related Work," we report the related work in robotics and we highlight our contribution more specifically. Then in Section "Proposed Method," we formulate the problem and our proposed solution. We provide details on the body schema implementation, on the robotic platform used and on the error metrics employed (see Experimental Setup) and we present experimental results in simulation and with the real robot (see Results). Then, in Section "Discussion," we discuss the proposed method and the results achieved. Finally, in Section "Conclusion and Future Work," we draw our conclusions and sketch the future work.

2. RELATED WORK

Reaching for objects and manipulating them is a crucial behavior in both humans and robots. While the classical approach in robotics is to rely on analytical models for motion control, humans learn such models from motor experience.

A number of works have proposed computational models to acquire these abilities through learning, without relying on any explicit model (Reinhart and Steil, 2009; Ciancio et al., 2011; Caligiore et al., 2014; Peniak and Cangelosi, 2014).

An alternative approach is to learn a model from sensorimotor data, and use the model for control. Such a model is typically referred to as "body schema."

The acquisition and adaptation of a robot body schema has been a topic of considerable attention [see, for example, Hoffmann et al. (2010) for a review up to the year 2010]. Learning (or adapting) the body schema of a humanoid robot can be seen also as a calibration problem, in which the goal is to align the reference frame located in the eyes, where visual information about the environment is obtained, with the one centered in the hand (i.e., eye-hand calibration). Clearly, in order to accurately

perform reaching and grasping actions a good calibration of these reference frames is required.

Since the visual estimation of the hand pose is a very challenging task, a way to simplify the calibration problem is to use a marker to visually detect the end-effector (i.e., the robot hand, in the humanoid case). For instance, the method used by Birbach et al. (2012) requires 5 min of data acquisition during specific robot movements with a special marker in the robot wrist. It optimizes offline some parameters of the kinematic chain (angle offsets and elasticity) of an upper humanoid torso using nonlinear least squares.

Online solutions have been studied, for example, in Ulbrich et al. (2009) and Jamone et al. (2012), in which visual markers are used to easily detect the hand position. The inclusion of additional parts into the kinematic chain (i.e., tools) has been considered as well in Jamone et al. (2013a,b).

In general, the adoption of a learning-by-doing strategy, in which a model is learned online during the execution of a goal-driven movement (goal-directed exploration (Jamone et al., 2011) or goal babbling (Rolf, 2013)), has been shown to improve learning performances, for example, by reducing the time required for convergence. Moreover, it allows to learn not only forward models but also inverse models, including, for example, the inverse kinematics of a redundant robot system (Rolf et al., 2010; Damas et al., 2013).

Although visual servoing techniques have been studied since the early 1980s (Agin, 1980) and a number of advanced solutions have been proposed during the last 30 years (Chaumette and Hutchinson, 2007), real-time reaching and grasping tasks in humanoid robots are often performed without any visual feedback of the hand (Saxena et al., 2008; Ciocarlie et al., 2010). Also according to Bohg et al. (2014), very few methods for grasping control take advantage of vision to correct the pose of the robot end-effector. The main reason for this is that the visual estimation of the pose of the end-effector is difficult to achieve and computationally expensive; therefore, such visual feedback is

typically noisy and cannot be obtained at a fast rate. However, purely open-loop reaching and grasping can hardly be successful, because the robot models are typically not accurate enough. For example, in Figueiredo et al. (2012) grasping is performed in kitchenware objects using a very precise robotic arm; however, some of the grasping experiments failed due to "contact between the hand and the object before grasping." These undesired and unexpected premature contacts can be mitigated with a visual servoing approach.

Chaumette and Hutchinson (2006) define visual servoing as a feedback closed-loop control strategy based on vision. Many visual servoing applications rely on eye-in-hand frameworks, where the cameras are attached to the robot end-effector (La Anh and Song, 2012; Ma et al., 2013). In humanoid robots, the cameras are placed in the head, and visual servoing can be done with an eye-to-hand approach (Hutchinson et al., 1996). Most applications of eye-to-hand visual servoing use markers in the end-effector in order to estimate its pose.

In Kulpate et al. (2005), the use of a single camera, a landmark in the hand (a light bulb emitting a red light) and a flat mirror, improves the estimation of the hand position and orientation. In Vahrenkamp et al. (2008), a red ball is attached to the robot wrist to allow for precise grasping using stereo calibrated vision. The humanoid REEM is used in Agravante et al. (2013) to perform reaching and grasping with visual feedback, using special markers on the hand and on the objects; the results show how the reaching motion planned on the basis of the robot kinematic model was not accurate enough to allow for precise object grasping, and how the inclusion of a visual servoing component could accommodate for such inaccuracies.

Marker-free solutions have been explored as well, either for body schema adaptation or for visual servoing control. The solution proposed in Ulbrich et al. (2012) is based on the decomposition of the kinematic chain into smaller segments; then, both offline and online learning solutions are proposed to learn the kinematic structure of the robot. Although it seems that no markers are used, no description is present about how the end-effector pose is measured. In Fanello et al. (2014), eye-hand calibration is realized by performing several ellipsoidal arm movements with a predefined hand posture, tracking the tip of the index finger in the camera images. Optimization techniques are employed to learn the transformation between the fingertip position obtained by the stereo vision and the one computed from the forward kinematics. Such transformation is then used to calibrate the kinematics. However, the hand orientation is not considered.

A marker-less visual servoing strategy can be used if one can estimate the pose of the robot hand using visual data. This is a challenging problem *per se* that has been studied both in Human-Computer Interaction and in Robotics (see Erol et al. (2007) for a review up to the year 2007). A few interesting works in robotics have used machine learning techniques to deal with the problem of robot hand detection. Leitner et al. (2013) used the Cartesian Genetic Programming method to learn how to detect the robot hand inside an image from visual examples. Online Multiple Instance Learning was used by Ciliberto et al. (2011) for the same task (detect the robot hand), through the use of

proprioceptive information from the arm joints and visual optic flow to automatically label the training images. However, in both works the hand orientation is neglected – only the position of the hand is learned. The work by Gratal et al. (2011) proposes a 3D-model based approach and an edge based error function to estimate the pose of the Schunk Dexterous Hand. This method is similar to ours as they exploit also graphics acceleration techniques and an edge-based approach. However, their optimization method is based on Virtual Visual Servoing (Comport et al., 2006) that, being a gradient based method, is prone to converge to local minima. On the contrary, we propose a sequential Monte Carlo method that is robust to non-convex/non-gaussian error functions.

2.1. Our Contribution

This paper extends our previous work on eye-hand adaptation in a humanoid robot (Vicente et al., 2014) and its GPGPU implementation (Vicente et al., 2015), by (i) comparing the influence of the number of particles in the estimation of the pose of the end-effector; (ii) performing an edge-based likelihood on the real robotic platform, and (iii) exploiting the derived models for the closed-loop control of the robot end-effector using visual feedback.

Our proposed system outperforms the related works described above by combining a number of features that are, in our opinion, fundamental to obtain an accurate control of goal-directed movements in humanoid robots. Our method does not use any special marker in the robot end-effector or in the robot wrist – it is a marker-free system. We estimate the 6D end-effector pose, rather than only the position, and we perform this estimation online during reaching tasks – our method does not require the execution of specific movements to calibrate the body schema. Moreover, we exploit the body schema adaptation and the real-time estimation of the end-effector pose to perform a marker-free visual servoing control strategy that improves the accuracy of reaching movements. To the best of our knowledge, a system that combines these fundamental features and that is successfully implemented in a real humanoid robot was not yet proposed in the literature.

3. PROPOSED METHOD

The body schema adaptation can be seen as an internal process that occurs in the mind of the robot and on the perception of the self. We focus on the perception of the arms and hands by the visual system placed in the robot head. This problem is known in human sciences as eye-hand coordination. In robotics, the eye-hand coordination relies on computing the transformation between two reference frames: (i) the eye reference frame and (ii) the end-effector reference frame. In our case, the first is located in the center of the left-eye and the latter in the center of the hand palm (see **Figure 2**). Moreover, the end-effector pose is defined as the pose of the hand palm in the eye reference frame. In this work, we estimate the end-effector pose using vision (left and right images) and proprioception (encoder readings), and adapt the initial body schema to reduce the mismatches between the internal
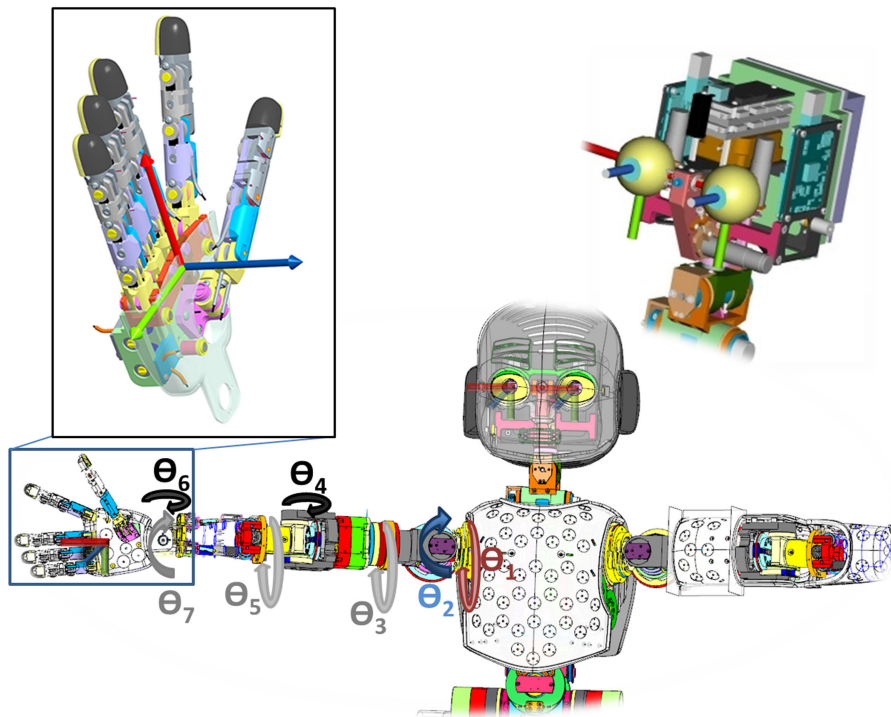
FIGURE 2 | The end-effector's pose (palm of the hand) is a function of the joints angles. Axis color notation: x-red; y-green; z-blue. (For a more detailed description visit: http://wiki.icub.org/wiki/ICubForwardKinematics). Best seen in color.

model prediction of the end-effector pose and the observed end-effector pose. According to the free-energy principle (Friston, 2010), biological agents try to minimize its free energy with the environment to achieve equilibrium. The free-energy principle tries to mathematically define how humans and animals optimize their expectation of the world. Thus, the free-energy measures the surprise present in perception given a generative model. The agent can suppress free energy by acting on the world (exciting the sensory input) or by changing its generative model to compensate the perception. Moreover, if we see the agent's body schema as the hidden state of the generative model, one of the solutions to achieve the equilibrium is to perform the body schema adaptation, as proposed in this work.

To achieve this goal, we consider two phases of reaching movements. First, an open-loop ballistic movement drives the end-effector to the vicinity of the target without visual feedback. During this period, vision is used to estimate the end-effector pose and adapt the internal model but the arm controller does not use this information. Second, a closed-loop control based on vision drives the robot's end-effector to the desired final pose, relative to the target of interest. During this stage, the internal model continues its adaptation based on vision and the arm controller used the adapted model to move the arm. In this section, we describe our methodology to address these phases. We begin by introducing the body schema model of our humanoid robot. Then, we explain the end-effector's vision based pose estimation method during the ballistic movement. Finally, we describe how to perform the control of the arm using the visual feedback.

3.1. Body Schema Modeling

Let us consider the problem of estimating the robot's end-effector pose in the left camera's reference frame (an analogous analysis can be done for the right camera). The real pose ($\mathbf{x}^r$) can be represented by a generic 4×4 roto-translation matrix $\mathbf{T}$. Using the robot kinematics function from the left camera to the hand palm $\mathscr{K}(\cdot)$ and the vector of joint encoder readings θ (see **Figure 2**), an estimate of the pose can be obtained by:

$$\hat{\mathbf{T}}_{\text{kin}} = \mathscr{K}(\theta) \tag{1}$$

However, several sources of error may affect this estimate. Let us consider the existence of calibration errors (bias). This source of error can be encoded in many different ways. We propose to encode it in the robot's joint space, i.e.,

$$\boldsymbol{\theta}^r = \boldsymbol{\theta} + \boldsymbol{\beta} \tag{2}$$

where $\boldsymbol{\theta}^r$ are the real angles; $\boldsymbol{\theta}$ are the measured angles; $\boldsymbol{\beta}$ are joint offsets representing calibration errors. Given an estimate of the joint offsets $\hat{\boldsymbol{\beta}}$, a better end-effector's pose estimate can be computed by:

$$\hat{\mathbf{T}}_{\text{joint}} = \mathscr{K}\left(\boldsymbol{\theta} + \hat{\boldsymbol{\beta}}\right). \tag{3}$$

Another solution is to encode the calibration error in Cartesian space using a roto-translation matrix defined as:

$$\hat{\mathbf{T}}_{\text{Cart}} = \mathscr{K}(\boldsymbol{\theta}) \cdot \hat{\mathbf{T}}_{\text{ERR}} \tag{4}$$

where $\mathbf{T}_{\text{ERR}}$ encodes the calibration errors.

The generalization of the learned parameters to other parts of the workspace was analyzed before in Vicente et al. (2015). We have shown that a parameterization of the error in the body schema in terms of joint offsets generalizes better to other parts of the workspace when compared to the non-calibrated case and to the Cartesian Error modeling, because the dominant sources of error are actually joint offsets.

Therefore, in this work, the learning process of the internal model consists in estimating the joint's offsets ($\boldsymbol{\beta}$) in the kinematic chain [see equation (2)]. Moreover, as we have access to the proprioceptive feedback ($\boldsymbol{\theta}$), estimating the joint offsets rather than the absolute joint values is a more effective approach: (i) the search space is smaller and (ii) we can use the adapted body schema (learned offsets) in other movements without re-learning it from scratch.

3.2. State Estimation with Sequential Monte Carlo Methods

Let $\mathbf{x}$ be a generic state vector and $\mathbf{y}$ the observation vector. Assuming that $\mathbf{y}$ depends stochastically on $\mathbf{x}$ at time t[1] one can devise a Bayesian filter to estimate the state $\mathbf{x}$ from the observation $\mathbf{y}$. The Bayesian filter consists of two steps: prediction and update. In the former, we calculate $\mathbf{x}_t$ from the previous state $\mathbf{x}_{t-1}$ according to the following equation:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1}) \cdot p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) \ d\mathbf{x}_{t-1} \quad (5)$$

where $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ is the transition probability and $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$ the previous estimation of the state at time $t-1$. In the second step, we update the posterior distribution with the last observation:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \eta \cdot p(\mathbf{y}_t|\mathbf{x}_t) \cdot p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \quad (6)$$

where η is a normalization factor and the probability $p(\mathbf{y}_t|\mathbf{x}_t)$ is called measurement probability.

The particle filter, also known as sequential Monte Carlo (SMC) method, is a non-parametric implementation of the Bayes filter, where we approximate the posterior distribution equation (6) by a finite number of samples, called particles:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \approx \sum_{m=1}^{M} \omega^{[m]} \delta(\mathbf{x}_t - \mathbf{x}_t^{[m]}) \cdot \left(\sum_{m=1}^{M} \omega^{[m]} \right)^{-1} \quad (7)$$

where M is the number of particles, $\mathbf{x}_t^{[m]}$ (with $1 < m < M$) is one particle, $\omega^{[m]}$ is the weight of particle m and $\sum_{m=1}^{M} \omega^{[m]} = 1$. The three stages of the particle filter are as follows:

1. Prediction: we sample $\mathbf{x}_t^{[m]}$ from $p(\mathbf{x}_t|\mathbf{x}_{t-1}^{[m]})$ adding these particles to a temporary set $\bar{\mathfrak{X}}_t$.
2. Update: we receive a new observation vector, $\mathbf{y}_t$, and update the particle weight or particle likelihood ($\omega^{[m]}$) according to:
$\omega^{[m]} = p(\mathbf{y}_t|\mathbf{x}_t^{[m]})$.

---

[1] In other words, $\mathbf{x}$ and $\mathbf{y}$ belong to a generative model also known as hidden Markov model.

3. Re-sampling: the particles are sampled according to their weight: $\omega^{[m]}$. This step is of paramount importance for the particle filter algorithm to work properly, Thrun et al. (2005) called it: the "trick" of the algorithm. We replace the M particles in the temporary set $\bar{\mathfrak{X}}_t$ by another M particles according to their weights $\omega^{[m]}$. Whereas in the temporary $\bar{\mathfrak{X}}_t$ the particles were distributed according to equation (5), after this step they are distributed (approximately) according to the posterior [see equation (6)].

For further details on Bayes and Particle filters, one can read Thrun et al. (2005).

3.3. Parameter Estimation with Sequential Monte Carlo Methods

In spite of being often used to track dynamic states, some modifications to the sequential Monte Carlo (SMC) methods have been proposed to estimate static parameters as well. In Kantas et al. (2009), the authors perform an overview of SMC methods for parameter estimation. Let, again, $\mathbf{x}$ be the initial non-static state vector and $\boldsymbol{\beta}$ the static parameter vector. An augmented state is defined as follows:

$$\mathbf{x}^{\text{aug}} = [\mathbf{x} \ \boldsymbol{\beta}]^T \quad (8)$$

One of the proposed solutions to estimate the parameters $\boldsymbol{\beta}$ is to introduce an artificial dynamics, changing from a static transition model:

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} \quad (9)$$

to a slowly time-varying one:

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \mathbf{w} \quad (10)$$

where $\mathbf{w}$ is an artificial dynamic noise that decreases when t increases.

3.3.1. Our Formulation

In our particular case, we are interested in the estimation of the end-effector pose ($\mathbf{x}$) as well as the calibration error parameters $\boldsymbol{\beta}$. We define the augmented state vector at time t as:

$$\mathbf{X}_t^{\text{aug}} = [\ \text{vec}(\mathcal{K}(\boldsymbol{\theta}_t + \boldsymbol{\beta}_t))\ \boldsymbol{\beta}_t]^T \quad (11)$$

where $\mathcal{K}(\cdot)$ is the robots kinematics function [see equation (3)] and the vector $\boldsymbol{\beta}$ is composed of the offsets in the kinematics chain from the camera to the end-effector. To reduce the complexity of the problem, we only consider the angular offsets of the arm kinematic chain (7 DOF), as the head chain is assumed to be calibrated, for instance, using the procedure defined in Moutinho et al. (2012). Also miscalibration in the finger joints has a smaller impact in the observations since they are at the end of the kinematic chain.

The offsets in equation (11) define the parameter vector, $\boldsymbol{\beta} = [\beta_1 \ \beta_2 \ \beta_3 \ \beta_4 \ \beta_5 \ \beta_6 \ \beta_7]^T$, as an unobserved Markov process where β_i is the offset in joint i of the arm assuming an initial distribution:

$$p(\boldsymbol{\beta}_0) \quad (12)$$

According to the general model in equation (10), $\boldsymbol{\beta}$ is the vector composed by the offsets in the arm, and the artificial noise

$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{K})$ is a zero mean Gaussian noise with a given diagonal covariance $\mathbf{K} = \sigma_s^2 \mathbf{I}_7$ and σ_s is an appropriately defined SD reflecting the magnitude of the calibration errors.

The first part of the augmented state, $\mathcal{H}(\boldsymbol{\theta}_t + \boldsymbol{\beta}_t)$, is deterministic given $\boldsymbol{\beta}_t$ since it is based on the robot kinematics function and the noiseless encoder readings, thus estimating the posterior distribution of the full state is equivalent to posterior distribution of the particles:

$$p(\boldsymbol{\beta}_t | \mathbf{y}_{1:t}, \boldsymbol{\theta}_{1:t}) \equiv p(\mathbf{x}_t^{\text{aug}} | \mathbf{y}_{1:t}) \tag{13}$$

We use a SMC method to approximate the posterior distribution defined in equation (13) by a set of random samples (particles):

$$\mathbf{B}_t := \left\{ \boldsymbol{\beta}_t^{[1]}, \boldsymbol{\beta}_t^{[2]}, \boldsymbol{\beta}_t^{[3]}, \ldots, \boldsymbol{\beta}_t^{[M]} \right\} \tag{14}$$

where M is the number of particles, $\boldsymbol{\beta}_t^{[m]}$ (with $1 < m < M$) is one state sample, and $\mathbf{B}_t$ is the particle set at time t. The *a posteriori* density distribution is approximated by the weighted set of particles:

$$p(\boldsymbol{\beta}_t | \mathbf{y}_{1:t}, \boldsymbol{\theta}_{1:t}) \approx \sum_{m=1}^{M} \omega^{[m]} \delta(\boldsymbol{\beta}_t - \boldsymbol{\beta}_t^{[m]}) \cdot \left(\sum_{m=1}^{M} \omega^{[m]} \right)^{-1} \tag{15}$$

where $\omega^{[m]}$ is the weight of particle m, $\delta(.)$ is the Dirac delta function, and the last factor is the normalization factor. In the beginning of each time step t, all the particles have the same weight: $\omega^{[m]} = \frac{1}{M}$. Under the Markov assumption, we can compute recursively $p(\boldsymbol{\beta}_t | \mathbf{y}_{1:t}, \boldsymbol{\theta}_{1:t})$ sampling from the previous estimation $p(\boldsymbol{\beta}_{t-1} | \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{1:t-1})$.

The filter has three stages as defined in Section "State estimation with Sequential Monte Carlo Methods": prediction, update, and re-sampling:

1. Prediction: we sample $\boldsymbol{\beta}_t^{[m]}$ from $p(\boldsymbol{\beta}_t | \boldsymbol{\beta}_{t-1}^{[m]})$, according to the transition [see equation (10)], in our case $\boldsymbol{\beta}_t^{[m]} = \boldsymbol{\beta}_{t-1}^{[m]} + \mathbf{w}$.

2. Update: we receive a new observation vector, $\mathbf{y}_t$, and update the particle weight ($\omega^{[m]}$) according to: $\omega^{[m]} = p(\mathbf{y}_t | \boldsymbol{\beta}_t^{[m]})$, i.e., the particle likelihood. Our observation model and the particle likelihood are defined in Section "Observation Model."

3. Re-sampling: the particles are sampled according to their weight using the systematic re-sampling method (Hol et al., 2006), which guarantees that a particle with a weight greater than $1/M$ is always re-sampled, where M is the total number of particles.

3.4. Observation Model

In this section, we address the problem of how to calculate the measurement probability in equation (6). The measurement probability can also be seen as the particle weight/likelihood after normalization. Our humanoid robot has two sources of information: (i) cameras on the eyes (visual sensing) and (ii) head and arm encoders (proprioceptive sensing). These two sources of information are related by the following model:

$$\mathbf{y}_t = \mathcal{F}(\boldsymbol{\theta}_t + \boldsymbol{\beta}_t) + \boldsymbol{\eta} \tag{16}$$

where $\boldsymbol{\theta}_t$ are the encoder readings and $\boldsymbol{\beta}_t$ the actual offsets in the joints at time step t. The function $\mathcal{F}(.)$ encodes the kinematic structure [see equation (1)], appearance of the robot, the camera's intrinsic parameters, and the image rendering model provided by a computer graphics engine able to generate realistic views of the robot. The actual observation, $\mathbf{y}_t$, is a random variable that concatenates the images acquired from the left and right cameras and $\boldsymbol{\eta}$ an image random noise (due to diverse non-modeled sources, e.g., specularities, shadows, camera jitter, etc., not necessarily Gaussian).

To sample from this model, we use the computer graphics rendering engine that generates virtual images of the robots cameras for arbitrary values of the vector $\boldsymbol{\beta}$ and encoder readings $\boldsymbol{\theta}$ (see **Figure 3**):

$$\hat{\mathbf{y}}_t^{[m]} = \mathcal{F}(\boldsymbol{\theta}_t + \boldsymbol{\beta}_t^{[m]}) \tag{17}$$

where $\hat{\mathbf{y}}_t^{[m]}$ represents the concatenation of the virtual images in the left and right cameras of the robot simulator for each generated hypothesis (particle). The particles can be seen as the multiple hypotheses generated by the brain while imagining the possible images consistent with the current state.

From the comparison between the real measurements $\mathbf{y}_t$ and the virtual measurements $\hat{\mathbf{y}}_t^{[m]}$, through a suitable function $g(\mathbf{y}_t, \hat{\mathbf{y}}_t^{[m]})$, we can compute the likelihood of $\boldsymbol{\beta}$ at time t:

$$l(\boldsymbol{\beta}_t^{[m]}) = p(\mathbf{y}_t | \boldsymbol{\theta}_t, \boldsymbol{\beta}_t^{[m]}) \propto g(\mathbf{y}_t, \hat{\mathbf{y}}_t^{[m]}) \tag{18}$$

We have defined two different approaches for implementing the comparison function $g(\cdot, \cdot)$. One is based on the hand's silhouette through image segmentation and the other is based on image contours through edge extraction.

3.4.1. Silhouette Segmentation

In this approach, we use the segmented binary images from the real and virtual cameras (see **Figure 4**). To compute the similarity between the real and virtual binary masks (silhouettes), we use the Jaccard coefficients (s_{Jc}) (see Cox and Cox (2000) for more detail). Let $\mathbf{R}(\mathbf{y})$ be the real silhouette and $\mathbf{R}(\hat{\mathbf{y}})$ the virtual silhouette. The Jaccard coefficient is defined as follows:

$$s_{Jc}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{\#\left(\mathbf{R}(\mathbf{y}) \cap \mathbf{R}(\hat{\mathbf{y}})\right)}{\#\left(\mathbf{R}(\mathbf{y}) \cup \mathbf{R}(\hat{\mathbf{y}})\right)} \tag{19}$$

where # denotes the number of pixels in the region.

The numerator term in equation (19) is measuring how similar and overlaid are the two silhouette regions and the denominator is normalizing the metric to a range [0,1]. Therefore, we define the likelihood model as follows:

$$p(\mathbf{y}_t | \boldsymbol{\beta}_t, \boldsymbol{\theta}_t) \propto s_{Jc}(\mathbf{y}_t, \hat{\mathbf{y}}_t) \tag{20}$$

In order to apply this approach, we need a good segmentation of the area of interest in the image. In this work, this is a feasible approach if one of the following conditions are met: either the head of the robot is static and a silhouette can be extracted by background segmentation methods, or the background is uniformly colored and a good silhouette can be obtained by color segmentation methods. In case, the head is moving or the background is cluttered, this approach is not robust and the edge-based method, described in the following section, is preferred.
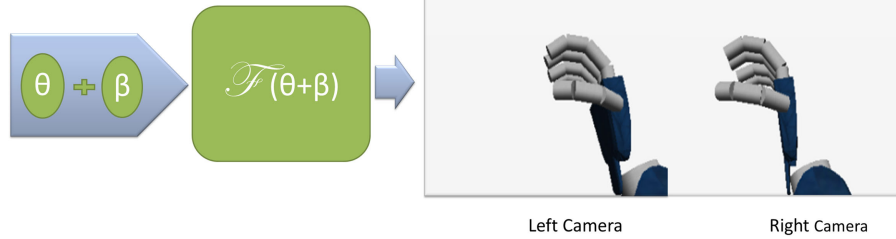
FIGURE 3 | Image generation in the internal mental simulator. The observation is defined as the concatenation of the left and right cameras.
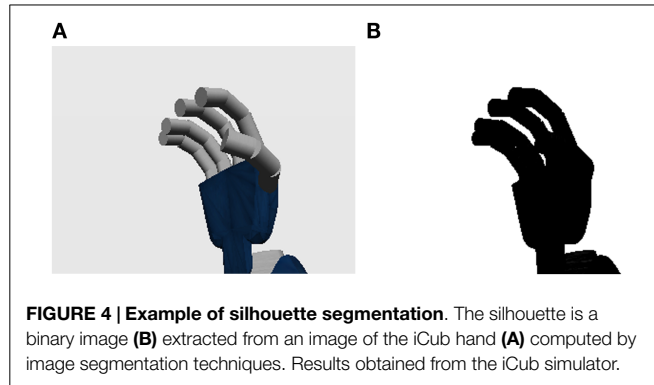

FIGURE 4 | Example of silhouette segmentation. The silhouette is a binary image **(B)** extracted from an image of the iCub hand **(A)** computed by image segmentation techniques. Results obtained from the iCub simulator.

3.4.2. Edge Extraction Approach

In this approach, the segmentation of the area of interest is not needed, instead, we exploit the edge information extracted from images, which is more robust to clutter, thus, more suitable to realistic environments. For this approach, we compute the average distance between the edges of the real image to the closest edge in the virtual image, and denote this quantity $\bar{d}$. A perfect match between the real and virtual images will correspond to $\bar{d} = 0$ whereas bad matches will correspond to large values of $\bar{d}$. The likelihood function is thus defined as:

$$p(\mathbf{y}_t|\boldsymbol{\beta}_t, \boldsymbol{\theta}_t) \propto \exp^{-\lambda_{\text{edge}} \cdot \bar{d}} \qquad (21)$$

where λ_{edge} is a tuning parameter to control sensitivity in the distance metric.

To compute $\bar{d}$, we make use of the distance transform (Borgefors, 1986). The distance transform (DT) consists in the application of an edge detector to the image (e.g., Canny (1986)) and then, for each pixel, compute its distance to the closest edge point. This distance has a minimum of 0 pixel and a maximum of 255 pixel, since the DT result is a 8-bit single-channel image. In **Figure 5**, we give an example of the right camera's image and the corresponding edge and distance transform images.

Let $\mathbf{D}(\mathbf{y})$ be the distance transform of the real images and $\mathbf{E}(\hat{\mathbf{y}}^{[m]})$ be the edge map of the virtual images (binary image indicating the edge pixels).

The average distance, $\bar{d}^{[m]}$ for each particle, can be efficiently computed using the Chamfer matching distance (Borgefors, 1988) defined as follows:

$$\bar{d}^{[m]} = \frac{1}{k} \cdot \sum_{i=0}^{N} \mathbf{E}\left(\hat{\mathbf{y}}^{[m]}(i)\right) \cdot \mathbf{D}\left(\mathbf{y}(i)\right) \qquad (22)$$

where k is the number of edge pixels in the virtual image, i is an index that runs over all pixels, and N is the total number of pixels.

3.5. Computing the Parameter Estimate

Although the parameters are represented at each time step as a distribution approximated by the particles, for practical purposes we must compute our best guess of the parameter vector $\hat{\boldsymbol{\beta}}$. We use a kernel density estimation (KDE) to smooth the weight of the particles according to the information of neighbor particles, and choose the particle with the highest smoothed weight ($\omega'^{[i]}$) as our parameter estimate:

$$\omega'^{[i]} = \omega^{[i]} + \alpha \cdot \frac{1}{M} \sum_{m=0}^{M} \omega^{[m]} \cdot K(\boldsymbol{\beta}^{[i]}, \boldsymbol{\beta}^{[m]}) \qquad (23)$$

where $\omega^{[i]}$ is the particle likelihood, α is a smoothing parameter, M is the number of particles used in our SMC implementation, and $\boldsymbol{\beta}^{[i]}$ is the particle we are smoothing. The sum term is the influence of the neighbors in the score of particle i. K is a kernel specifying the influence of one particle in others based on their distance. We use a Gaussian Kernel in our experiments:

$$K(\boldsymbol{\beta}^{[i]}, \boldsymbol{\beta}^{[j]}) = \frac{1}{\sqrt{2\pi|\Sigma|}} e^{[-\frac{1}{2}(\boldsymbol{\beta}^{[i]}-\boldsymbol{\beta}^{[j]})^T \Sigma^{-1}(\boldsymbol{\beta}^{[i]}-\boldsymbol{\beta}^{[j]})]} \qquad (24)$$

where Σ is the co-variance matrix and $|\Sigma|$ its determinant.

Since the joints' offsets ($\boldsymbol{\beta}$) are independent of each other, Σ will be a diagonal matrix:

$$\Sigma = \sigma_{\text{KDE}}^2 \cdot I_7 \qquad (25)$$

where σ_{KDE} is the SD in each joint, which we assume equal. This parameter determines if two particles are close or not. If we have a high σ_{KDE}, all particles will be "close" to each other. On the other hand, if we have a small σ_{KDE}, all particles will be fairly isolated in the world resulting in $\omega'^{[i]} \approx \omega^{[i]}$.

It is worth to note that due to the redundancy in the robot kinematics (joints space is 7DOF while the end-effector pose is 6DOF) different solutions in the set B_t may correspond to the same target pose. Therefore, the likelihood function $l(\boldsymbol{\beta})$ is multimodal and a particular choice of $\hat{\boldsymbol{\beta}}$ will be just one set of offsets that can explain the end-effector's appearance in the images. For this reason, the proposed method with sequential Monte Carlo, which does not assume any particular distribution of the posterior, is a suitable parameter estimation approach.
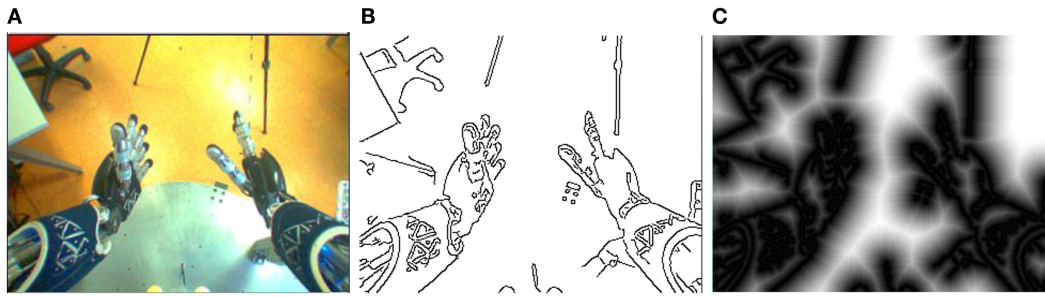
FIGURE 5 | Example of the computation of the edges and distance transform of the iCub hand in a real environment on the right camera. (A similar example can be shown for the left camera.) **(A)** is the input image, **(B)** shows the edges extraction using Canny, 1986, and **(C)** the distance transform using Borgefors, 1986.

3.6. Controlling the End-Effector

In this work, we focus on the control of the arm to obtain a desired end-effector's pose using the robot internal body schema. We have implemented two control modalities: (i) an open-loop "ballistic" movement and (ii) a closed-loop strategy exploiting visual feedback.

3.6.1. Open-Loop

The open-loop control is the dominant control mode in robotics manipulation. It relies on accurate calibration of the robot system and accurate object sensing. It exploits the inverse kinematics of the head-arm-hand chain, from the eye to the end-effector and only uses visual sensing for the initial estimation of the object/target pose. During arm control, only proprioceptive feedback is used. The open-loop control relies on solving the robot's inverse kinematics ($\mathscr{K}^{-1}$):

$$\mathbf{q}^{\mathrm{d}} = \mathscr{K}^{-1}(\mathbf{x}^{\mathrm{d}}) \qquad (26)$$

where $\mathbf{q}^d$ is the joints configuration (command) that leads to the desired end-effector's pose ($\mathbf{x}^d$) and $\mathscr{K}^{-1}$ the inverse kinematics function.

The trajectory between the initial joints configuration ($\mathbf{q}^i$) and the desired one ($\mathbf{q}^d$) is a linear trajectory in the joint space, performing a movement with a constant velocity according to the following equation:

$$\mathbf{q}_{t+1} = \mathbf{q}_t + \frac{\mathbf{q}^d - \mathbf{q}^i}{\Delta t} \qquad (27)$$

where $\mathbf{q}_t$ is the joint command at time t and Δt is the desired movement duration.

3.6.2. Closed-Loop

In the closed-loop approach, instead of controlling the position of the joints, we control the joint velocities $\dot{\mathbf{q}}$ based on visual feedback.

As mentioned in Section "Related Work," our problem is an instance of eye-to-hand visual servoing. Two control modalities are common in visual servoing approaches: (i) image-based control, where the arm's motion is determined by the error between the current and desired configurations in the image coordinates

or (ii) a position-based approach, where the arm's motion is determined by the error between the current and desired 6D poses of the end-effector. Our approach is a position-based strategy in an eye-to-hand configuration. See Hutchinson et al. (1996) for a more detailed taxonomy of visual servoing strategies.

Following the notation in Siciliano and Khatib (2007), the error ($\mathbf{e}$) to be minimized is defined as follows:

$$\mathbf{e} = \mathbf{x}^{\mathrm{c}} - \mathbf{x}^{\mathrm{d}} \qquad (28)$$

where $\mathbf{x}^{\mathrm{d}}$ is the desired 6D pose of the end-effector and $\mathbf{x}^{\mathrm{c}}$ the current.

The relationship between the joint velocities and the time variation of the 6D pose error is given by:

$$\dot{\mathbf{e}} = \mathbf{J}(\mathbf{q}) \cdot \dot{\mathbf{q}} \qquad (29)$$

where $\mathbf{J}(\mathbf{q})$ is the robot Jacobian from the left-eye to the end-effector reference frame.

If we defined $\dot{\mathbf{e}} = -\lambda \cdot \mathbf{e}$ (to ensure an exponential decoupled decreasing error) and invert the robot Jacobian ($\mathbf{J}(\mathbf{q})$) by using the Moore-Penrose pseudo inverse, we end up with the following control law:

$$\dot{\mathbf{q}} = -\lambda \cdot \mathbf{J}^{\dagger}(\mathbf{q}) \cdot \mathbf{e} \qquad (30)$$

where $\mathbf{J}^{\dagger}(\mathbf{q})$ is the pseudo-inverse in the joint angles $\mathbf{q}$.

In our case, as we correct the joint angles on the robot arm, the control law with the improved robot Jacobian will be:

$$\dot{\mathbf{q}} = -\lambda \cdot \mathbf{J}^{\dagger}(\boldsymbol{\theta} + \hat{\boldsymbol{\beta}}) \cdot \mathbf{e} \qquad (31)$$

where $\boldsymbol{\theta}$ are the encoder readings and $\hat{\boldsymbol{\beta}}$ the joint offset vector.

4. EXPERIMENTAL SETUP

4.1. Robotic Platform

The iCub (see **Figure 1**) is a humanoid robot for research in artificial intelligence and cognition. It has 53 motors that move the legs, waist, head, arms, and hands, and it has the average size of a 3-year-old child. It was developed in the context of the EU project RobotCub (2004–2010) and subsequently adopted by more than 25 laboratories worldwide. Its stereo vision system (cameras in the

eyeballs), proprioception (motor encoders), touch (tactile finger-tips and artificial skin), and vestibular sensing (IMU on top of the head) are important characteristics that allow the study of autonomy in humanoid robots. The robot is equipped with a dynamic simulator (Tikhanoff et al., 2008) that can be controlled using the same software that is used for the real robot. We resort to this simulator in a number of experiments in order to evaluate the performance of our approach with precise ground-truth (that we cannot access on the real robot).

4.2. Body Schema

The body schema can be considered the agent's knowledge about the kinematics, posture, and appearance of its body parts. The body schema is a mental state that includes sensory information about the self and the world and about the relationships between the body parts. In this work, we have implemented the iCub's body schema on 3D computer graphics engines. Graphics engines permit an effective generation of mental images of body states through the knowledge of the kinematic structure of the robot, and its body appearance. The internal mental simulator projects the 3D simulated body into the robot vision. In particular, we are interested in the projection of the arms and hands – the end-effectors. The capabilities of the internal mental simulator are similar to the real agent: (i) we can control the end-effector to a given pose, (ii) it has proprioceptive sensing, thus, we can acquire the current joint values of the arm and head, and (iii) it has stereo vision – projecting the 3D world into 2D images.

4.3. Error Metrics

4.3.1. Position and Orientation

In order to evaluate the accuracy of our method, we compute the Cartesian error ($E_{\text{Cartesian}}$) composed of position and orientation errors between two generic poses, A and B, as:

$$E_{\text{Cartesian}} = [d_o, d_p] \qquad (32)$$

The general orientation error (d_o) is defined as:

$$d_o(\mathbf{R}_A, \mathbf{R}_B) = \sqrt{\frac{|| \log m(\mathbf{R}_A^T \mathbf{R}_B)||_F^2}{2}} \frac{180}{\pi} \quad [°] \qquad (33)$$

where $\mathbf{R}_A$ and $\mathbf{R}_B$ are two rotation matrices from the eye reference frame to the end-effector frame. The principal matrix logarithm, logm, with the Frobenius norm, ($||\cdot||_F$), implements the usual distance on the group of rotations. The general position error between the two different poses is computed by the Euclidean distance, $d\text{p}(\mathbf{P}_A, \mathbf{P}_B)$:

$$d_p(\mathbf{P}_A, \mathbf{P}_B) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2} \qquad (34)$$

where $\mathbf{P}_A$ and $\mathbf{P}_B$ are 3D Cartesian positions of the end-effector.

4.3.2. Defined Poses

In this work, we define four different poses. The real pose ($\mathbf{x}^r$) is defined as:

$$\mathbf{x}^r = [\mathbf{P}^r \ \text{vec}(\mathbf{R}^r)] \qquad (35)$$

where $\mathbf{R}^r$ is the real rotation matrix and $\mathbf{P}^r$ the real 3D position. This pose is the ground-truth data for evaluating the method. In the simulation experiments, this is the pose with the introduced artificial offsets. The second pose is the desired pose ($\mathbf{x}^d$) which is the pose that we want to achieve during the reaching task:

$$\mathbf{x}^d = [\mathbf{P}^d \ \text{vec}(\mathbf{R}^d)] \qquad (36)$$

The initial pose ($\mathbf{x}^i$) is the initial joint configuration at the beginning of the reaching movement. Finally, the estimated pose ($\mathbf{x}^e$) that is the robot's forward kinematics applied to the sum of the measured joint angles $\boldsymbol{\theta}$ (the proprioception) and the estimated $\boldsymbol{\beta}$ (or $\boldsymbol{\beta} = 0$ when the adaptation is not performed):

$$\mathbf{x}^e = [\mathbf{P}^e \ \text{vec}(\mathbf{R}^e)] \qquad (37)$$

4.3.3. Estimation and Reaching errors

The estimation error – $E_{\text{estimation}}$ – is the difference between the real pose ($\mathbf{x}^r$) and the estimated pose ($\mathbf{x}^e$) using equations (33) and (34):

$$E_{\text{estimation}} = [d_o(\mathbf{R}^r, \mathbf{R}^e) \ , \ d_p(\mathbf{P}^r, \mathbf{P}^e)] \qquad (38)$$

The real reaching error – E_{reaching}^r – is defined as the difference between the desired ($\mathbf{x}^d$) and real pose ($\mathbf{x}^r$):

$$E_{\text{reaching}}^r = [d_o(\mathbf{R}^d, \mathbf{R}^r) \ , \ d_p(\mathbf{P}^d, \mathbf{P}^r)] \qquad (39)$$

It measures how far the end-effector is from the target pose.

The estimated reaching error – E_{reaching}^e – is defined as the difference between the desired ($\mathbf{x}^d$) and estimated pose ($\mathbf{x}^e$):

$$E_{\text{reaching}}^e = [d_o(\mathbf{R}^d, \mathbf{R}^e) \ , \ d_p(\mathbf{P}^d, \mathbf{P}^e)]. \qquad (40)$$

It represents the robot's belief on how far its end-effector is from the target pose.

4.4. Computer Specifications

The experiments were performed in a computer equipped with an Intel® Xeon® Processor W3503 at 2.4 GHz with two cores, two threads, and a 4-MB memory cache and a NVidia GeForce GTX 750 with 512 CUDA Cores, a base clock of 1020 MHz and 2048 MB of memory (RAM).

4.5. Experimental Settings

In this section, we describe the experimental parameters, common to all the presented results. We initialize the SMC with $M = 200$ particles, defining $p(\boldsymbol{\beta}_0) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ [see equation (12)] with a given diagonal covariance $\mathbf{Q} = \sigma_i^2 \mathbf{I}_7$ and $\sigma_t = 5°$. In all the experiments we started from scratch, i.e., the best estimation at $t = 0$ is the proprioception of the robot ($\boldsymbol{\beta} = 0$). The artificial dynamic noise is initialized with $\sigma_s = 4°$ and it decreases with t by a factor of 0.8:

$$\sigma_s(t) = \sigma_s(t - 1) * 0.8 \qquad (41)$$

where t is the frame index. This value has a lower bound of 0.08° to allow continuous adaptation.

The kernel density estimation was initialized using a SD of $\sigma_{\text{KDE}} = 1°$ in equation (23) and a neighborhood influence of $\alpha = 500$ in equation (25).

5. RESULTS

In this section, we report the experimental results. We divide them into two parts: simulations (Section "Simulation Results") and real-world evaluations (Section "Real Robot Results"). In the former, we evaluate quantitatively our method comparing the body schema adaptation with the ground-truth measurements and in the latter we test our approach qualitatively in the real robot.

In Section "Reaching Movement and Body Schema Adaptation," we show how the online adaptation of the body schema and the estimation of the pose of the end-effector allow to accurately reach for a desired pose, using a combined open-loop and closed-loop control scheme: (i) during the open-loop control (described in Section "Open-Loop"), the online body schema adaptation is performed, allowing for better estimation of the end-effector pose and (ii) during the closed-loop control (described in Section "Closed-Loop"), the end-effector pose feedback is exploited to accurately reach for the desired pose.

Then, in Section "Trade-Off between the Number of Particles and Estimation Accuracy," we assess the performance of the internal model estimation procedure, evaluating the relationship between the number of particles used in the optimization procedure and the accuracy of the estimation.

Finally, in Section "Real Robot Results," we show the method working in the real world: the iCub robot performs online adaptation of its body schema (i.e., including both arms) and real-time estimation of the pose of its end-effectors (i.e., both right and left hands) exploiting visual feedback from its cameras, in an unstructured environment (i.e., with natural background in the images).

5.1. Simulation Results

In the simulation experiments, we use the iCub simulator both as robot and as internal mental simulation. The iCub simulator is a realistic software that uses ODE (Open Dynamic Engine) for simulating the motion of rigid bodies and their physical interaction. It uses the same software and control architecture of the real iCub robot. In order to consider the iCub simulator a realistic model of the real robot, we introduce artificial angular offsets in the 7 DOFs of the right arm kinematic chain. We define ETA = [5, 4, 3, −2, 3, −7, 3]; these offsets have the same order of magnitude of the calibration errors, we typically encounter on the real robot. We use the same set of offsets in all the simulation experiments, in order to be able to compare the different results. Therefore, in these experiments the only difference between the robot and the internal mental simulation is the set of artificial offset; the goal of the body schema online adaptation is to compensate for these offsets.

Hand visual perception relies on the silhouette segmentation approach (described in Section "Silhouette Segmentation"); a homogeneous white background is located in front of the robot and the segmentation is performed based on color information. In general, the silhouette approach is effective in cases where the segmentation is easy (e.g., with the white background). In Section "Real Robot Results," we will motivate the use of a different strategy, the edge-based approach (described in Section "Edge Extraction Approach"), for the real robot experiments, where we deal with natural background; such strategy could not, however,

be used in these simulation experiments, because the texture model of iCub simulator is poor (based on simple cylinders and cubes of a homogeneous gray color) and too few edges are present in the images.

5.1.1. Reaching Movement and Body Schema Adaptation

In this first set of experiments, we have two main objectives: (i) evaluate the error in the end-effector pose estimation equation (38) during the movements, and (ii) show the convergence of the reaching error (real and estimated, equations (39) and (40), respectively) during the closed-loop control made possible by the body schema adaptation.

We define a constant duration of the open-loop phase (120 frames) in order to estimate a stable solution for the joint offsets (β) and we define 50 frames in the close-loop as the maximum number of frames to acquire during the reaching to the desired pose $\mathbf{x}^d$ using visual feedback. The error decaying factor presented in the closed-loop control section was initialized with the value $\lambda = 5$ [see equation (31)].

Overall, we perform 120 movements with different initial and final poses in order cover different areas of the working space: 4 different final poses with 10 initial poses, with 3 repetition of each movement. The results in this section show the mean and SD over the 120 experiments performed. We initialize our sequential Monte Carlo implementation with 200 particles and with $\beta = 0$, i.e., we always start the movements with the nominal non-calibrated model.

In **Figure 6**, we can see the mean and the SD of the end-effector pose estimation error during the movements. We show the error both with the nominal non-calibrated model (without online adaptation) and with the online adaptation. The algorithm converges to a good estimation after about 60 frames, improving it during the last part of the movement. It can be noticed that, in spite of having constant artificial offsets in the joints (i.e., a constant source of error), the pose estimation error in the non-calibrated case (without online adaptation, red dotted line in **Figure 6**) is not constant and it depends on the current arm configuration.

In **Figure 7**, the evolution of the reaching error [see equation (39)] during the whole movement (open-loop and closed-loop) is displayed. Both the mean and the variance of the error over the 40 movements are shown. The high variance at the beginning of the motion is due to different (10) initial poses of the end-effector; some of them are closer (~60 mm) to the target pose than others (~120 mm). During the reaching movement, this variance reduces as the arm goes to the different (4) target poses; the variance is due to different arm configuration with constant artificial joint offsets.

The open-loop part of the movement is planned at the movement onset, based on the non-calibrated model. Therefore, the reaching error at the end of the open-loop phase is equal to the pose estimation error with the non-calibrated system (as it can be seen by comparing the red dotted lines in **Figure 6**) to the line in **Figure 7** at frame 120, for both position and orientation. Then, the body schema adaptation performed during the open-loop phase can be exploited at the onset of the closed-loop phase to obtain an accurate estimate of the pose of the end-effector; this allows to consistently reduce the reaching error already at frame 130.
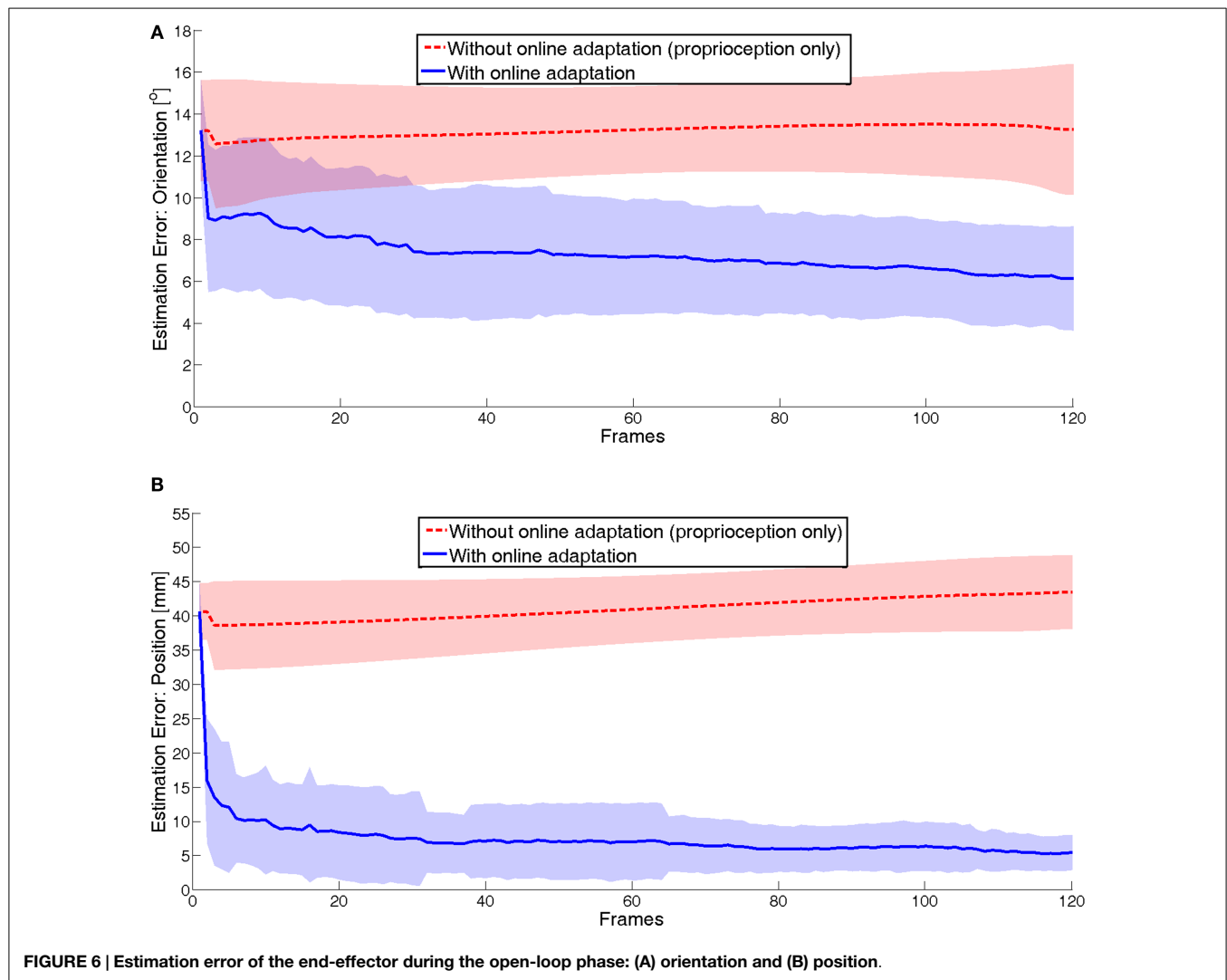
FIGURE 6 | Estimation error of the end-effector during the open-loop phase: (A) orientation and (B) position.

However, as expected, the error does not converge to 0 because there is still a residual error in the estimation of the end-effector pose (that can be appreciated in the blue bold line of the plot in **Figure 6**). **Figure 8** shows a close-up of the final frames of the movement. The estimated reaching error (the dotted green line) converges indeed to 0, indicating that the closed-loop control is working properly. However, as mentioned above, the real reaching error (black solid line, same information as in **Figure 7**) does not converge, due to the residual estimation errors.

Table 1 reports the exact numerical data related to the plots in **Figures 6** and **7**: the pose estimation error at the end of the open-loop phase (both with and without online adaptation) and the reaching error (both at the end of the open-loop and at the end of the closed-loop).

5.1.2. Trade-Off between the Number of Particles and Estimation Accuracy

To generate particles/images and to compare them online with the ones obtained from visual feedback requires a lot of computational effort in the two processing units (Central Processing Unit (CPU) and Graphical Processing Unit (GPU)); therefore, the overall

computational burden increases with the number of particles used in the SMC method. In order to better understand how the number of particles influences the accuracy of the estimation, we performed an extensive evaluation in which several movements are executed and the body schema adaptation is performed using different amounts of particles: $M = 100; 200; 500; 1000; 2000$. For each value of M, we perform 40 different movements with different initial and final poses. In each experiment, we maintain the parameters defined before in Section "Experimental settings" and we change only the amount of particles used; we performed the same motions of the right arm with the same visual feedback. The final end-effector pose estimation errors (mean and SD over the 40 movements) for each value of M are shown in **Figure 9**; then, in **Table 2** we report only the mean values, and we compare them to the non-adaptation case as well. A clear trend can be noticed, which relates the increase of the number of particles to the decrease of the estimation error. However, this relation is non-linear: the slope of the curve is higher in the beginning and lower in the end. Indeed, the difference between the use of $M = 1000$ and $M = 2000$ is quite small, which suggests that further increasing the number of particles would not improve
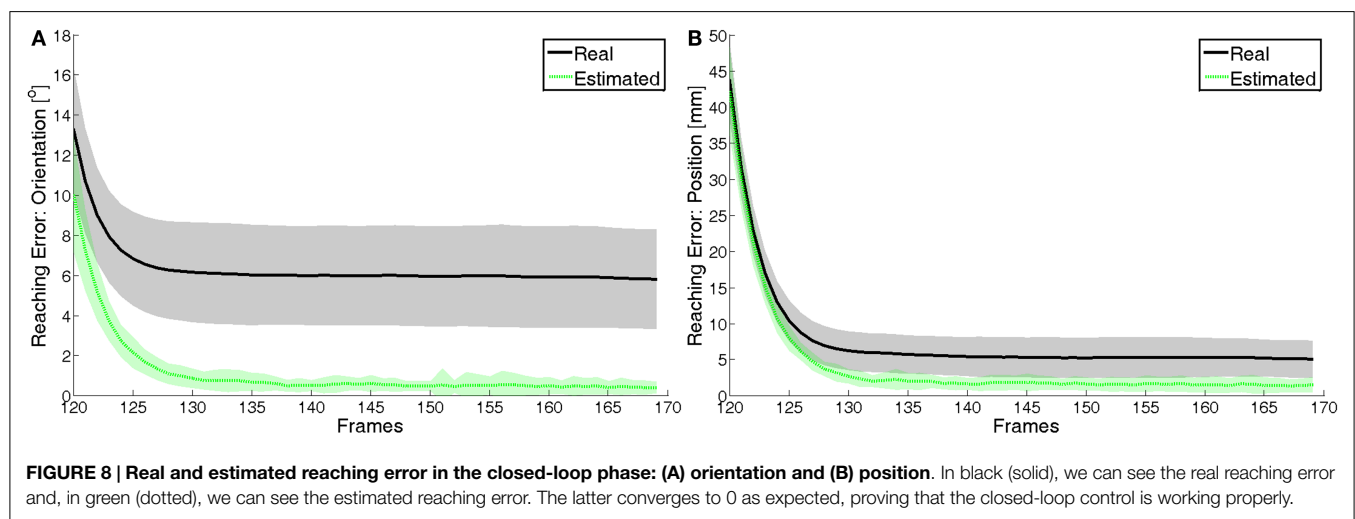
FIGURE 7 | Real reaching error during the whole movement: (A) orientation and (B) position. The error is the average over 120 experiments: 4 different final poses, 10 different initial poses and each movement is repeated three times. It can be seen how the closed-loop correction considerably reduces the reaching error.

FIGURE 8 | Real and estimated reaching error in the closed-loop phase: (A) orientation and (B) position. In black (solid), we can see the real reaching error and, in green (dotted), we can see the estimated reaching error. The latter converges to 0 as expected, proving that the closed-loop control is working properly.
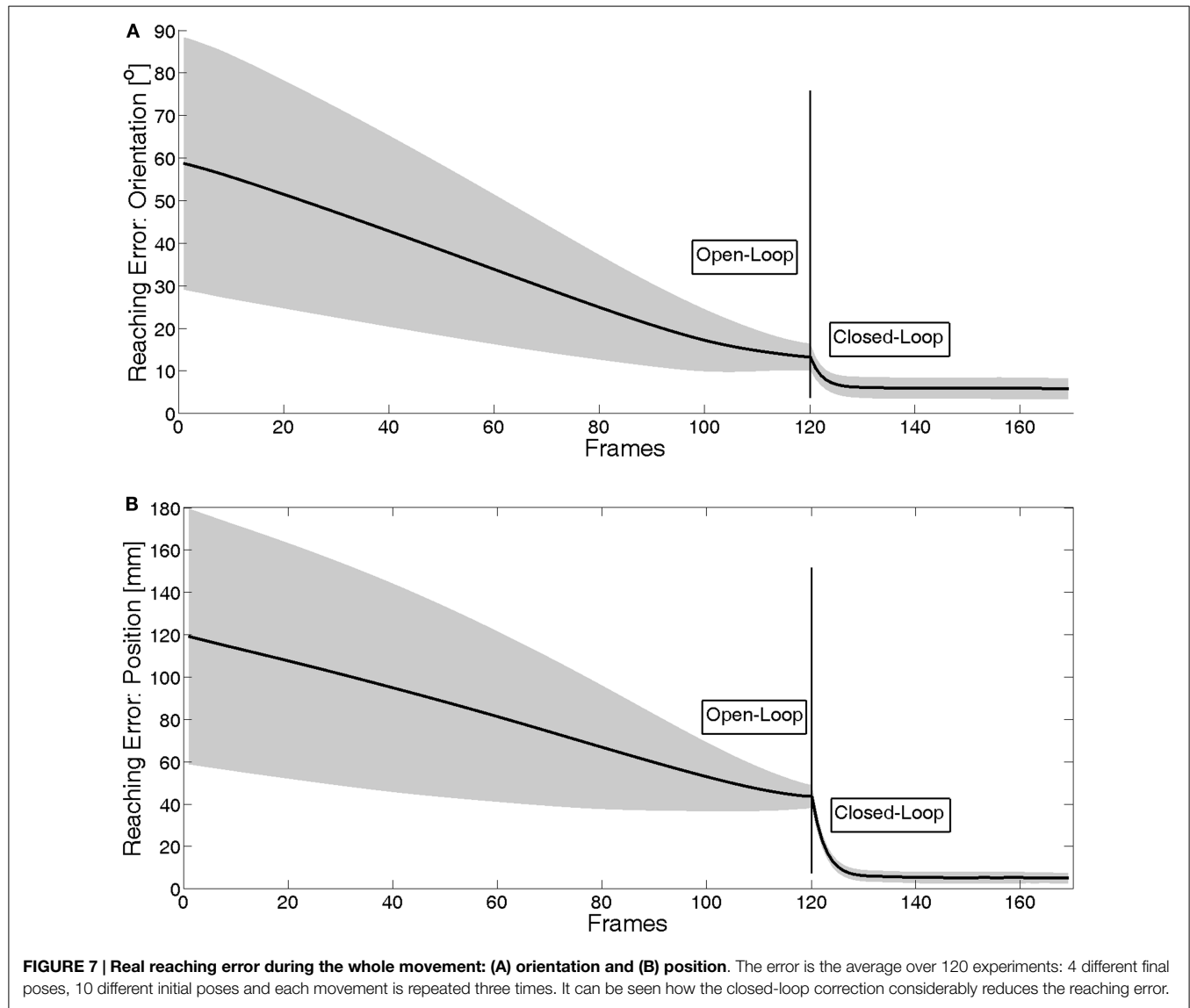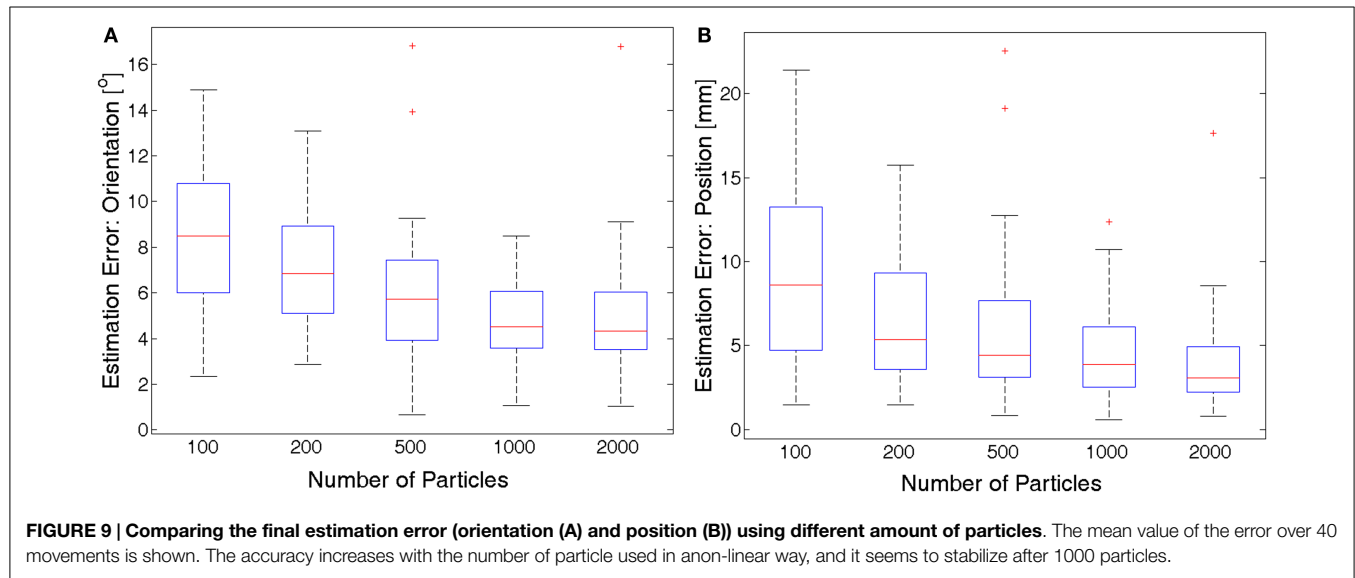
TABLE 1 | Estimation and real reaching errors in the final poses of each movement: the estimation errors and the real reaching errors of open-loop are computed at frame 120, in the final pose of the open-loop phase.

	Estimation error				Real reaching error			
	Without adaptation		With adaptation		Open-loop		Closed-loop	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Orientation error [°]	13.30	3.13	6.14	2.49	13.30	3.15	5.81	2.49
Position error [mm]	43.79	5.40	5.45	2.55	43.79	5.42	5.13	2.57

FIGURE 9 | Comparing the final estimation error (orientation (A) and position (B)) using different amount of particles. The mean value of the error over 40 movements is shown. The accuracy increases with the number of particle used in anon-linear way, and it seems to stabilize after 1000 particles.

TABLE 2 | Trade-off between number of particles and accuracy of the estimation.

	No adaptation	100 particles	200 particles	500 particles	1000 particles	2000 particles
Orientation error [°]	14.50	8.48	6.85	5.73	4.50	4.34
Position error [mm]	41.29	8.61	5.35	4.40	3.87	3.09

For each value of the number of particles, we show the average value of the estimation errors over the 40 test movements.

the estimation considerably. Moreover, it can be noticed that the estimation of the end-effector orientation benefits more of the increasing number of particles than the estimation of the end-effector position; this might be an indication that the orientation is more difficult to estimate.

In **Figure 10**, we show the temporal evolution of the pose estimation error during the arm movement in two representative cases: with $M = 200$ and $M = 2000$ particles. Although the orientation and position errors are smaller in the 2000 particles case, more computation is required with respect to the case with 200 particles (computation takes about 10 times longer). The time needed to generate and evaluate 200 particles is around 0.8 s per frame, while for 2000 particles is 7.5 s per frame. In other words, more time is needed to generate and evaluate the hypotheses and, therefore, the movement must be slower if we want to acquire the same number of frames/images.
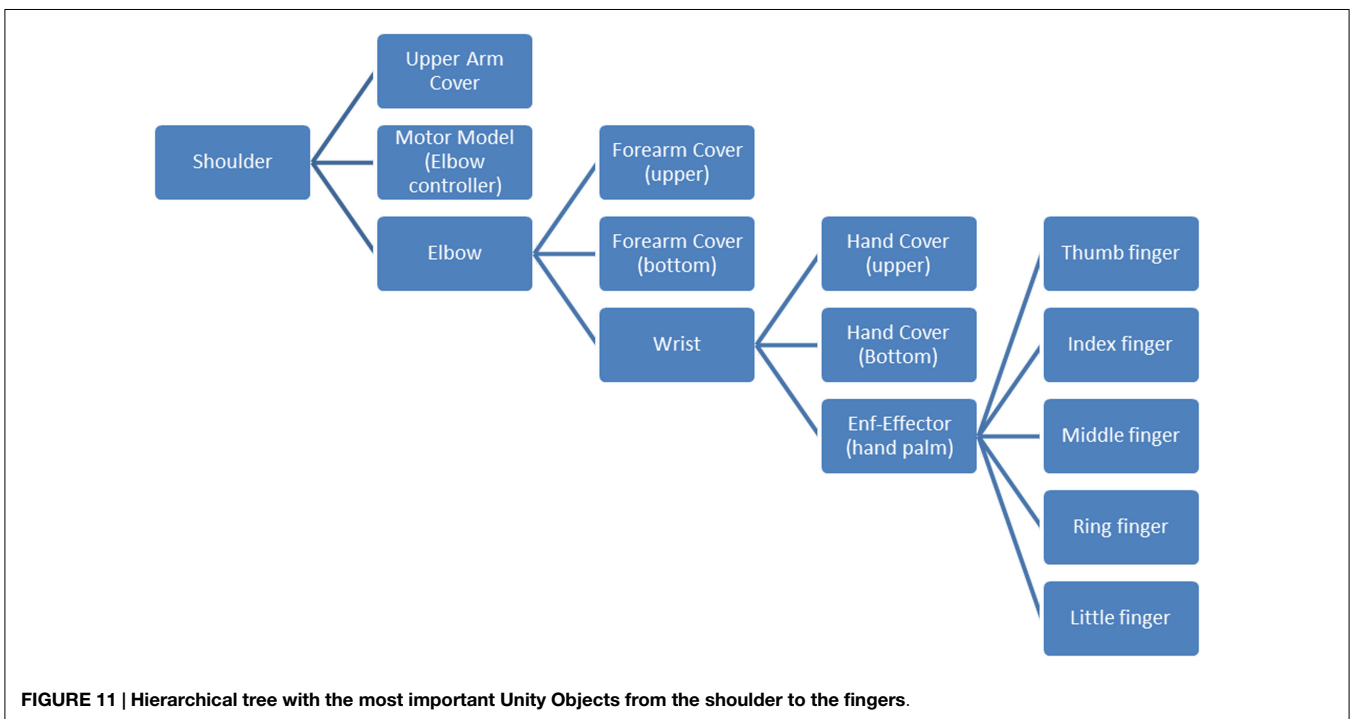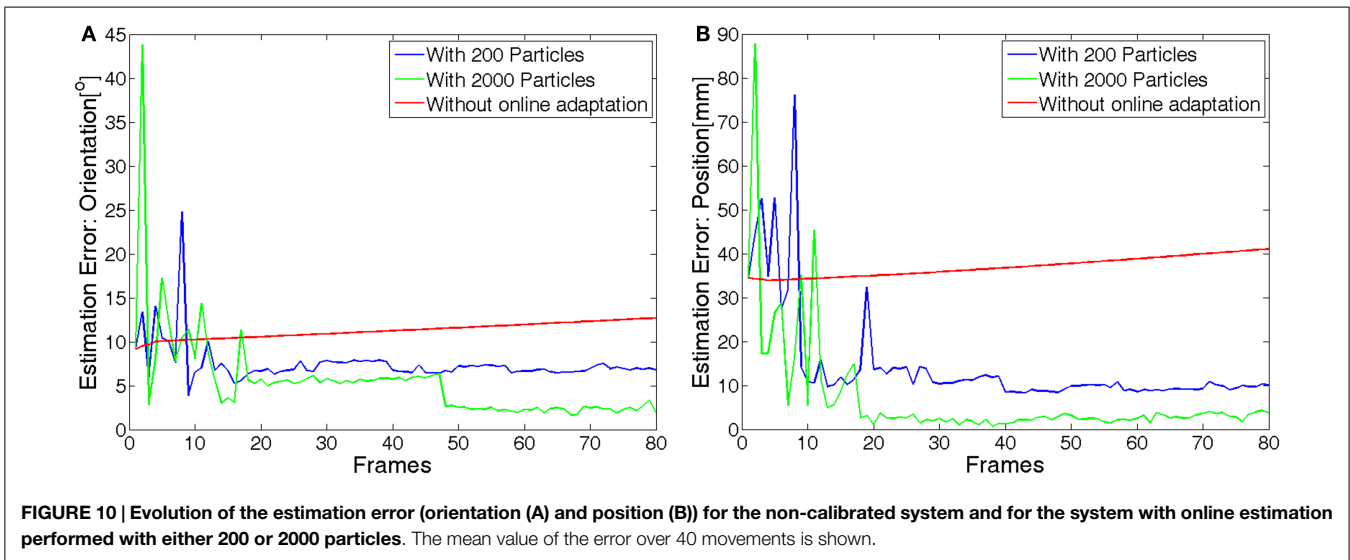
In summary, there is a trade-off between the accuracy of the estimation and the computation time for each iteration. In order to be able to perform the end-effector pose estimation in real-time

in our current computer system, we chose to use $M = 200$ particles; this choice allows us to perform the estimation during reaching movements performed at natural speed (i.e., in the order of 0.01 m/s measured on the end-effector), with an average estimation error of about 5.35 mm in position and 6.85° in orientation.

5.2. Real Robot Results

In the real-world experiments, we use the real iCub as robot (see **Figure 1**) and a Unity® computer graphics model as internal mental simulation.

Unity® is a renowned cross-platform game engine developed by Unity Technologies that can generate very realistic virtual images. While the iCub Simulator uses simplified meshes of the robot external surfaces, our Unity model of the iCub renders the full CAD model of the robot, thus providing a much better match of the real robot appearance; in particular, for our experiments, a good appearance model of the robot hands is crucial. To perform the internal mental simulation process in real-time, we rely on GPU programing to achieve faster computation, as described in

FIGURE 10 | Evolution of the estimation error (orientation (A) and position (B)) for the non-calibrated system and for the system with online estimation performed with either 200 or 2000 particles. The mean value of the error over 40 movements is shown.

FIGURE 11 | Hierarchical tree with the most important Unity Objects from the shoulder to the fingers.

details in Vicente et al. (2015). Moreover, at each time instant we render only the robot parts that are visible by the robot cameras using a shader in the graphics pipeline, instead of rendering the whole robot appearance. A hierarchical tree of the robot kinematics is defined where each node has a reference frame attached and a pivot point that is used to perform the rotation of this hierarchical object structure (See **Figure 11**). In other words, this tree represents the relationship between the several objects in the model (i.e., the robot body parts). For instance, the fingers are coupled with the robot hand, so that if the hand moves, the fingers will move along with it and update their absolute position in the world, maintaining the relative pose in the hand reference frame.

In these experiments, we exploit the edge-based approach for the hand visual perception, described in Section "Edge extraction approach." The silhouette approach that we used in the simulation experiments is not suitable in real-world scenarios due to the non-homogeneous background in the images, which makes segmentation difficult and noisy.

We maintain the initialization parameters defined in Section "Experimental Settings" and we define the tuning parameter of the edge distance as $\lambda_{edge} = 0.01$. This results in a higher likelihood when the distance of the nearest edge is around 1 pixel and a likelihood close to 0 when the distance reaches its maximum value (255).

As a way to evaluate the performance of the method in a real environment, we have chosen to use the left and right hands and apply the end-effector pose estimation on both. The goal is to get the hands close to each other with the index fingers almost touching. To achieve that the desired poses of the left and right end-effectors (i.e., the left and right hand palms) are defined with the same orientations, and with positions that differ only in the X-axis, of 12 cm; the fingers are slightly bent, so that the fingertips would touch when the hands are facing each other 12 cm apart. The target pose was chosen to be close to the center of both cameras for visualization purposes; however, this method can be applied in every location of the robot workspace, as long as the hands are in the field of view of one of the cameras. Therefore, the results shown are not specific to this target pose and similar experiments can be performed in any other configuration.

The robot starts from the home position seen in **Figure 12**. The left arm moves to the desired pose x^d and the target pose

for the end-effector of the right arm is defined to have the same orientation of the left arm end-effector and a distance in the perpendicular direction of the palm of 12 cm.

In the first part of the experiment, we control both the left and the right arms to the desired end-effector poses, with open-loop control, performing the body schema adaptation and the estimation of the poses of the end-effectors. In the second part of the experiment, we control the pose of both end-effectors to the desired poses with closed-loop control, exploiting the adapted body schema and the improved pose estimation. In **Figure 13**, we show the comparison between the non-calibrated case (after the open-loop control, top row), where the hands have a distance from each other of approximately 16 cm and the fingertips are not aligned, and the adaptation performed using our method (with the closed-loop control, bottom row), where the hands are about 12 cm apart and the fingertips are touching each other (as desired).
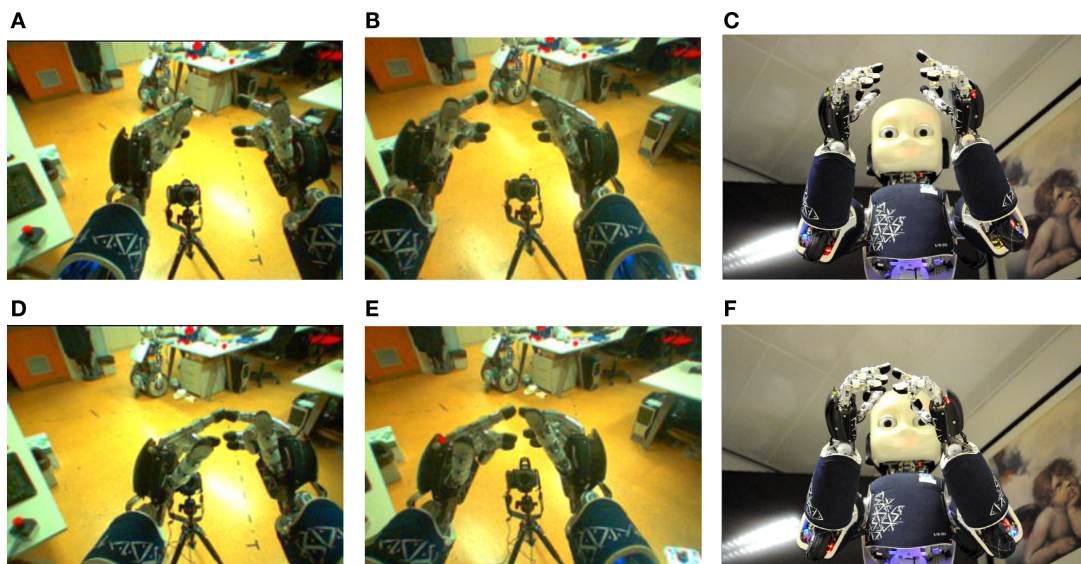

FIGURE 12 | Body schema online adaptation performed in the real robot. Images seen by the robot eye cameras Left **(A)** and Right **(B)** Cameras and by an external camera **(C)** placed in front of the robot. Initial robot configuration.

FIGURE 13 | Body schema online adaptation performed in the real robot. Images seen by the robot eye cameras Left **(A,D)** and Right **(B,E)** Cameras and by external camera **(C,F)** placed in front of the robot. First row **(A–C)**: the left arm is controlled toward a target end-effector pose and the right arm is controlled toward the same end-effector pose with a shift of 12 cm, with open-loop control. However, the resulting end-effector poses are not the desired one, due to inaccuracies in the body schema. Adaptation parameters are estimated during the motion of both arms, and used to update the body schema. Second row **(D–F)**: the pose of both end-effectors is corrected using the updated body schema and a closed-loop control strategy that exploits the improved pose estimation.

6. DISCUSSION

We have reported results both in simulation and in the real robot. The former constitute a quantitative evaluation with respect to ground-truth data; the latter demonstrate that the system can be used in the real world successfully. Indeed, in both cases, we achieve good results showing that both the estimation and reaching errors are decreased. Our approach is biologically inspired as evidence in neuroscience suggests that the human brain keeps an updated representation of the body (i.e., a body schema) that is employed to generate hypotheses of the limbs positions in space, which are combined with the actual perception of the self in a Bayesian fashion. Our system outperforms the current state of the art in the sense that (i) it does not rely on markers on the end-effector, using the pure visual feedback coming from the robot stereo cameras, (ii) the body schema adaptation is performed during reaching movements without a specific adaptation procedure, and (iii) such adaptation is performed online in real-time. While the body schema adaptation and pose estimation can be performed during each reaching movement, the adaptation parameters obtained during one movement generalize well to other areas of the workspace. This has been extensively documented in a previous publication (Vicente et al., 2015), in which we also show that our choice of parameterization (i.e., offsets in the arm joints) outperforms other solutions proposed in the literature, such as the use of offsets in the Cartesian position and orientation of the end-effector.

In general, the scalability of systems depends on the size of the search space (i.e., the parameters space). The fact that we parameterize the model with the joint offsets does not mean that other sources of error could not be accounted for. In theory, with a sufficient number of particles (and with a sufficient number of examples) any kind of error that causes a mismatch between the kinematic model and the real robot (e.g., unalignment of one joint axis of rotation, change in the length of one link, change in the elastic properties of one transmission cable) can be compensated for, since our sequential Monte Carlo parameter optimization approach attempts to minimize the prediction error between the body schema hypothesis and the visual perception. Although a quantitative analysis of the estimation with different error sources was not performed in this paper, the encouraging results obtained on the real robot (where other error sources than joint offsets are likely to be present) suggest that our system could deal with them.

The results provided in Section "Trade-Off between the Number of Particles and Estimation Accuracy" show that increasing the number of particles would lead to better estimation performance; however, the computational burden would also increase considerably. Interestingly, our architecture for the internal mental simulation could be easily made parallel to increase the computation speed. In the current system, one computer generates multiple hypotheses based on a single internal model; the number of generated hypothesis is the same of the number of particles. The hypothesis is then compared to the robot visual perception. The use of a big cluster of computers in which each machine runs an instance of the internal model and generates only a single hypothesis would considerably reduce the computation time, allowing to use a high number of particles (at the cost of using a high number of computers).

Our proposed solution is not robot-dependent, and can be applied to other robotic platforms in a straightforward manner, provided that a kinematic and graphical (texture) model of the robot is available. Clearly, the more the texture model of the robot is close to the real robot appearance, the better the estimation performance is expected to be. This is because in our current solution the appearance of the internal model is not updated exploiting the visual information gathered by the robot: only the kinematic structure is adapted based on the mismatch between the internal model predictions and the visual feedback.

7. CONCLUSION AND FUTURE WORK

We presented a novel system for simultaneous online body schema adaptation and end-effector pose estimation implemented on the iCub humanoid robot. The parameter adaptation is performed with a sequential Monte Carlo framework during the execution of reaching movements. We rely only on the robot embedded sensors (vision sensing from stereo cameras and proprioception) without using any special visual marker. Our method draws inspiration from human perception and learning, as we combine the prediction made by a learned internal model with the actual visual feedback to improve the perceptual skill of the robot.

Overall, our simulation experiments show that we can reduce the end-effector pose estimation error considerably with respect to using the nominal (non-calibrated) robot model (of about eight times in the end-effector position and 2.2 times in the end-effector orientation). Moreover, the use of a closed-loop correction after the initial open-loop reaching motion (during which the body schema adaptation and pose estimation are performed) allows to reduce the reaching error of about 8.5 times in position and 2.3 times in orientation.

We demonstrated the applicability of our system to real-world scenarios by performing a bimanual reaching task with the real iCub robot, where the combined open-loop and closed-loop control strategy, made possible by the accurate pose estimation, allowed to decrease the positioning error of both end-effectors by 4 cm.

Some possible directions for the future work have been discussed in Section "Discussion." Moreover, an interesting improvement to increase the robustness of the edge matching would be to use also the orientation of the matching edge on the model and compare its location and orientation with an edge in the realistic platform sensing information.

AUTHOR CONTRIBUTIONS

In this work, all the authors contributed to the conception of an online body schema adaptation solution and to the analysis and interpretation of the data acquired.

ACKNOWLEDGMENTS

REFERENCES

Agin, G. (1980). Computer vision systems for industrial inspection and assembly. *IEEE Comput.* 13, 11–20. doi:10.1109/MC.1980.1653613

Agravante, D., Pages, J., and Chaumette, F. (2013). "Visual servoing for the REEM humanoid robot's upper body," in *IEEE-RAS International Conference on Robotics and Automation (ICRA)* (Karlsruhe: IEEE), 5253–5258.

Ashmead, D., McCarty, M., Lucas, L., and Belvedere, M. (1993). Visual guidance in infants' reaching toward suddenly displaced targets. *Child Dev.* 64, 1111–1127. doi:10.1111/j.1467-8624.1993.tb04190.x

Berlucchi, G., and Aglioti, S. (1997). The body in the brain: neural bases of corporeal awareness. *Trends Neurosci.* 20, 560–564. doi:10.1007/s00221-009-1970-7

Birbach, O., Bäuml, B., and Frese, U. (2012). "Automatic and self-contained calibration of a multi-sensorial humanoid's upper body," in *IEEE-RAS International Conference on Robotics and Automation (ICRA)* (Saint Paul, MN: IEEE), 3103–3108.

Bohg, J., Morales, A., Asfour, T., and Kragic, D. (2014). Data-driven grasp synthesis – a survey. *IEEE Trans. Robot.* 30, 289–309. doi:10.1109/TRO.2013.2289018

Borgefors, G. (1986). Distance transformations in digital images. *Comput. Vis. Graph. Image Process.* 34, 344–371. doi:10.1016/S0734-189X(86)80047-0

Borgefors, G. (1988). Hierarchical chamfer matching: a parametric edge matching algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.* 10, 849–865. doi:10.1109/34.9107

Bushnell, E. (1985). The decline of visually guided reaching during infancy. *Infant Behav. Dev.* 8, 139–155. doi:10.1016/S0163-6383(85)80002-3

Caligiore, D., Parisi, D., and Baldassarre, G. (2014). Integrating reinforcement learning, equilibrium points, and minimum variance to understand the development of reaching: a computational model. *Psychol. Rev.* 121, 389–421. doi:10.1037/a0037016

Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 8, 679–698. doi:10.1109/TPAMI.1986.4767851

Chaumette, F., and Hutchinson, S. (2006). Visual servo control, part I: basic approaches. *IEEE Robot. Autom. Mag.* 13, 82–90. doi:10.1109/MRA.2006.250573

Chaumette, F., and Hutchinson, S. (2007). Visual servo control, part II: advanced approaches. *IEEE Robot. Autom. Mag.* 14, 109–118. doi:10.1109/MRA.2007.339609

Ciancio, A., Zollo, L., Guglielmelli, E., Caligiore, D., and Baldassarre, G. (2011). "Hierarchical reinforcement learning and central pattern generators for modeling the development of rhythmic manipulation skills," in *IEEE International Conference on Development and Learning (ICDL)*, Vol. 2 (Frankfurt: IEEE), 1–8.

Ciliberto, C., Smeraldi, F., Natale, L., and Metta, G. (2011). "Online multiple instance learning applied to hand detection in a humanoid robot," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (San Francisco, CA: IEEE), 1526–1532.

Ciocarlie, M., Hsiao, K., Jones, E. G., Chitta, S., Rusu, R. B., and Sucan, I. A. (2010). "Towards reliable grasping and manipulation in household environments," in *Internacional Symposium on Experimental Robotics (ISER)* (New Delhi: Springer), 241–252.

Comport, A., Marchand, E., Pressigout, M., and Chaumette, F. (2006). Real-time markerless tracking for augmented reality: the virtual visual servoing framework. *IEEE Trans. Vis. Comput. Graph* 12, 615–628. doi:10.1109/TVCG.2006.78

Cox, T. F., and Cox, M. (2000). *Multidimensional Scaling*, 2 Edn. New York: Chapman and Hall/CRC.

Damas, B., Jamone, L., and Santos-Victor, J. (2013). "Open and closed-loop task space trajectory control of redundant robots using learned models," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Tokyo: IEEE), 163–169.

Erol, A., Bebis, G., Nicolescu, M., Boyle, R. D., and Twombly, X. (2007). Vision-based hand pose estimation: a review. *Comput. Vis. Image Understand.* 108, 52–73. doi:10.1016/j.cviu.2006.10.012

Fanello, S. R., Pattacini, U., Gori, I., Tikhanoff, V., Randazzo, M., Roncone, A., et al. (2014). "3D stereo estimation and fully automated learning of eye-hand coordination in humanoid robots," in *IEEE-RAS International Conference on Humanoid Robots* (Madrid: IEEE), 1028–1035.

Figueiredo, R., Shukla, A., Aragao, D., Moreno, P., Bernardino, A., Santos-Victor, J., et al. (2012). "Reaching and grasping kitchenware objects," in *IEEE/SICE International Symposium on System Integration (SII)* (Fukuoka: IEEE), 865–870.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 2, 127–138. doi:10.1038/nrn2787

Gratal, X., Romero, J., and Kragic, D. (2011). "Virtual visual servoing for real-time robot pose estimation," in *18th World Congress of the International Federation of Automatic Control* (Milano: IFAC), 9017–9022.

Hoffmann, M., Marques, H., Hernandez Arieta, A., Sumioka, H., Lungarella, M., and Pfeifer, R. (2010). Body schema in robotics: a review. *IEEE Trans. Autonom. Ment. Dev.* 2, 304–324. doi:10.1109/TAMD.2010.2086454

Hol, J., Schon, T., and Gustafsson, F. (2006). "On resampling algorithms for particle filters," in *IEEE Nonlinear Statistical Signal Processing Workshop* (Cambridge: IEEE), 79–82.

Hutchinson, S., Hager, G., and Corke, P. (1996). A tutorial on visual servo control. *IEEE Trans. Robot. Autom.* 12, 651–670. doi:10.1109/70.538972

Jamone, L., Damas, B., Endo, N., Santos-Victor, J., and Takanishi, A. (2013a). Incremental development of multiple tool models for robotic reaching through autonomous exploration. *PALADYN J. Behav. Robot.* 03, 113–127. doi:10.2478/s13230-013-0102-z

Jamone, L., Damas, B., Santos-Victor, J., and Takanishi, A. (2013b). "Online learning of humanoid robot kinematics under switching tools contexts," in *IEEE-RAS International Conference on Robotics and Automation (ICRA)* (Karlsruhe: IEEE), 4811–4817.

Jamone, L., Natale, L., Hashimoto, K., Sandini, G., and Takanishi, A. (2011). "Learning task space control through goal directed exploration," in *IEEE International Conference on Robotics and Biomimetics (ROBIO)* (Phuket: IEEE), 702–708.

Jamone, L., Natale, L., Nori, F., Metta, G., and Sandini, G. (2012). Autonomous online learning of reaching behavior in a humanoid robot. *Int. J. HR* 09, 1250017. doi:10.1142/S021984361250017X

Joseph, R. (2000). Fetal brain behavior and cognitive development. *Dev. Rev.* 20, 81–98. doi:10.1006/drev.1999.0486

Kantas, N., Doucet, A., Singh, S., and Maciejowski, J. (2009). "An overview of sequential monte carlo methods for parameter estimation on general state space models," in *IFAC Symposium on System Identification (SYSID)* (Saint Malo: IFAC), 774–785.

Körding, K. P., and Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature* 427, 244–247. doi:10.1038/nature02169

Kulpate, C., Mehrandezh, M., and Paranjape, R. (2005). "An eye-to-hand visual servoing structure for 3d positioning of a robotic arm using one camera and a flat mirror," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Alberta: IEEE), 1464–1470.

La Anh, T., and Song, J.-B. (2012). Robotic grasping based on efficient tracking and visual servoing using local feature descriptors. *Int. J. Precis. Eng. Manuf.* 13, 387–393. doi:10.1007/s12541-012-0049-8

Leitner, J., Harding, S., Frank, M., Forster, A., and Schmidhuber, J. (2013). "Humanoid learns to detect its own hands," in *IEEE Congress on Evolutionary Computation (CEC)* (Cancun: IEEE), 1411–1418.

Lockman, J. J., Ashmead, D. H., and Bushnell, E. W. (1984). The development of anticipatory hand orientation during infancy. *J. Exp. Child Psychol.* 37, 176–186. doi:10.1016/0022-0965(84)90065-1

Ma, G., Huang, Q., Yu, Z., Chen, X., Meng, L., Sultan, M., et al. (2013). "Hand-eye servo and flexible control of an anthropomorphic arm," in *IEEE International Conference on Robotics and Biomimetics (ROBIO)* (Shenzhen: IEEE), 1432–1437.

Mathew, A., and Cook, M. (1990). The control of reaching movements by young infants. *Child Dev.* 61, 1238–1257. doi:10.2307/1130891

Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., et al. (2010). The icub humanoid robot: an open-systems platform for research in cognitive development. *Neural Netw.* 23, 1125–1134. doi:10.1007/978-3-540-77296-5_32

Miall, R. C., and Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Netw.* 9, 1265–1279. doi:10.1016/S0893-6080(96)00035-4

Moutinho, N., Brandao, M., Ferreira, R., Gaspar, J., Bernardino, A., Takanishi, A., et al. (2012). "Online calibration of a humanoid robot head from relative encoders, imu readings and visual data," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Vilamoura: IEEE), 2070–2075.

Pattacini, U. (2011). *Modular Cartesian Controllers for Humanoid Robots: Design and Implementation on the iCub*. Ph.D. thesis, Italian Institute of Technology, Genoa.

Peniak, M., and Cangelosi, A. (2014). "Scaling-up action learning neuro-controllers with GPUs," in *International Joint Conference on Neural Networks (IJCNN)* (Beijing: IEEE), 2519–2524.

Reinhart, R., and Steil, J. (2009). "Reaching movement generation with a recurrent neural network based on learning inverse kinematics for the humanoid

robot icub," in *9th IEEE-RAS International Conference on Humanoid Robots (Humanoids)* (Paris: IEEE), 323–330.

Rochat, P. (1998). Self-perception and action in infancy. *Exp. Brain Res.* 123, 102–109. doi:10.1007/s002210050550

Rolf, M. (2013). "Goal babbling with unknown ranges: a direction-sampling approach," in *IEEE Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL-Epirob)* (Osaka: IEEE), 1–7.

Rolf, M., Steil, J., and Gienger, M. (2010). Goal babbling permits direct learning of inverse kinematics. *IEEE Trans. Autonom. Ment. Dev.* 2, 216–229. doi:10.1109/TAMD.2010.2062511

Saxena, A., Driemeyer, J., and Ng, A. Y. (2008). Robotic grasping of novel objects using vision. *Int. J. Rob. Res.* 27, 157–173. doi:10.1177/0278364907087172

Siciliano, B., and Khatib, O. (2007). *Springer Handbook of Robotics.* New York: Springer-Verlag.

Sober, S. J., and Sabes, P. N. (2005). Flexible strategies for sensory integration during motor planning. *Nat. Neurosci.* 8, 490–497. doi:10.1038/nn1427

Thrun, S., Burgard, W., and Fox, D. (2005). *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents).* Cambridge: The MIT Press.

Tikhanoff, V., Fitzpatrick, P., Nori, F., Natale, L., Metta, G., and Cangelosi, A. (2008). "The icub humanoid robot simulator," in *IROS Workshop on Robot Simulators* (Nice).

Ulbrich, S., de Angulo, V., Asfour, T., Torras, C., and Dillmann, R. (2009). "Rapid learning of humanoid body schemas with kinematic bézier maps," in *IEEE-RAS International Conference on Humanoid Robots (Humanoids)* (Paris: IEEE), 431–438.

Ulbrich, S., de Angulo, V. R., Asfour, T., Torras, C., and Dillmann, R. (2012). General robot kinematics decomposition without intermediate markers. *IEEE Trans. Neural Netw. Learn. Syst.* 23, 620–630. doi:10.1109/TNNLS.2012.2183886

Vahrenkamp, N., Wieland, S., Azad, P., Gonzalez, D., Asfour, T., and Dillmann, R. (2008). "Visual servoing for humanoid grasping and manipulation tasks,"

in *IEEE-RAS International Conference on Humanoid Robots* (Daejeon: IEEE), 406–412.

Van der Meer, A. (1997). Keeping the arm in the limelight: advanced visual control of arm movements in neonates. *Eur. J. Paediatr. Neurol.* 1, 103–108. doi:10.1016/S1090-3798(97)80040-2

Vicente, P., Ferreira, R., Jamone, L., and Bernardino, A. (2014). "Eye-hand online adaptation during reaching tasks in a humanoid robot," in *IEEE International Conferences on Development and Learning and Epigenetic Robotics (ICDL-Epirob)* (Genoa: IEEE), 175–180.

Vicente, P., Ferreira, R., Jamone, L., and Bernardino, A. (2015). "Gpu-enabled particle based optimization for robotic-hand pose estimation and self-calibration," in *IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)* (Vila Real: IEEE), 3–8.

von Hofsten, C. (1991). Structuring of early reaching movements: a longitudinal study. *J. Mot. Behav.* 23, 280–292. doi:10.1080/00222895.1991.9942039

von Hofsten, C. (2004). An action perspective on motor development. *Trends Cogn. Sci.* 8, 266–272. doi:10.1016/j.tics.2004.04.002

Wolpert, D. M., Miall, R. C., and Kawato, M. (1998). Internal models in the cerebellum. *Trends Cogn. Sci.* 2, 338–347. doi:10.1007/978-94-007-1333-8_56

APPENDIX

Pseudo-code

In this appendix, we will give details on the implementations of some of the modules used in our approach. We show the pseudo-code of the most important modules developed.

We developed two modules: main module (**Algorithm 1**) and the internal mental simulation (**Algorithm 2**). In our work, they communicate via YARP middle-ware.

The main module is responsible for processing the images from the robot to generate the particles and updated their likelihood. Moreover, it publishes the estimated β that can be used to correct the end-effector pose. In the internal mental simulator, we generate the hypotheses and test them returning the likelihood of each particle.

ALGORITHM 1 | Main module.

---

procedure MAIN
    for $t \leftarrow 0$, MaxFrames **do**   ▷ e.g., MaxFrames $= 170$
        Receive images from the robot cameras
        Receive Joint encoders (θ)
        $y_t =$ Process-Images(RC, LC)
        $\mathbf{B}_t =$ SMC($\mathbf{B}_{t-1}$, $\mathbf{y}_t$, θ)
        Perform Kernel Density Estimation on set $\mathbf{B}_t$
        Publish estimation β_t
    end for
end procedure

procedure PROCESS-IMAGES(RC,LC)
    if Silhouette-based **then**
        Silhouette segmentation
        Concatenation of Images – $\mathbf{y}_t$
    end if
    if Edge-based **then**
        Edge Extraction   ▷ using (Canny, 1986)
        Distance transform   ▷ using (Borgefors, 1986)
        Concatenation of Images – $\mathbf{y}_t$
    end if
    return $\mathbf{y}_t$
end procedure

procedure SMC($\mathbf{B}_{t-1}$, $\mathbf{y}_t$, θ)
    $\hat{B}_t = B_t = \emptyset$
    for $m \leftarrow 1$, M **do**   ▷ M is the number of particles
        sample $\beta_t^{[m]} \sim p(\beta_t | \beta_{t-1}^{[m]})$
        $w_t^{[m]} =$ Simulator($\beta_t^{[m]}$, $\mathbf{y}_t$, θ)   ▷ According to $w_t^{[m]} = p(y_t | \beta_t^{[m]})$
        $\hat{B}_t = \hat{B}_t + \langle \beta_t^{[m]}, w_t^{[m]} \rangle$
    end for
    for $m \leftarrow 1$, M **do**   ▷ Re-sampling Stage
        draw $\beta_t^{[m]}$ with probability $\propto w_t^{[m]}$
        add $\beta_t^{[m]}$ to $\mathbf{B}_t$
    end for
    return $\mathbf{B}_t$
end procedure

ALGORITHM 2 | Internal mental simulator.

---

procedure SIMULATOR($\beta_t^{[m]}$, $\mathbf{y}_t$, θ)
    Generate pose with $\beta_t^{[m]}$ and θ
    if Silhouette-based **then**
        Generate binary Images $\hat{\mathbf{y}}_t$
        $w_t^{[m]} =$ Likelihood-Assessment($\hat{\mathbf{y}}_t$, $\mathbf{y}_t$)   ▷ Using GPGPU programing
    end if
    if Edge-based **then**
        Generate Edge Images $\hat{\mathbf{y}}_t$
        $w_t^{[m]} =$ Likelihood-Assessment($\hat{\mathbf{y}}_t$, $\mathbf{y}_t$)   ▷ Using GPGPU programing
    end if
    return $w_t^{[m]}$
end procedure

Finding Your Way from the Bed to the Kitchen: Reenacting and Recombining Sensorimotor Episodes Learned from Human Demonstration

Erik A. Billing[1]*, Henrik Svensson[1], Robert Lowe[1,2] and Tom Ziemke[1,3]

[1] Interaction Lab, School of Informatics, University of Skövde, Skövde, Sweden, [2] Interaction, Cognition and Emotion Lab, Department of Applied IT, University of Gothenburg, Gothenburg, Sweden, [3] Cognition and Interaction Lab, Department of Computer and Information Science, Linköping University, Linköping, Sweden

Several simulation theories have been proposed as an explanation for how humans and other agents internalize an "inner world" that allows them to simulate interactions with the external real world – prospectively and retrospectively. Such internal simulation of interaction with the environment has been argued to be a key mechanism behind mentalizing and planning. In the present work, we study internal simulations in a robot acting in a simulated human environment. A model of sensory–motor interactions with the environment is generated from human demonstrations and tested on a Robosoft Kompaï robot. The model is used as a controller for the robot, reproducing the demonstrated behavior. Information from several different demonstrations is mixed, allowing the robot to produce novel paths through the environment, toward a goal specified by top-down contextual information. The robot model is also used in a covert mode, where the execution of actions is inhibited and perceptions are generated by a forward model. As a result, the robot generates an internal simulation of the sensory–motor interactions with the environment. Similar to the overt mode, the model is able to reproduce the demonstrated behavior as internal simulations. When experiences from several demonstrations are combined with a top-down goal signal, the system produces internal simulations of novel paths through the environment. These results can be understood as the robot imagining an "inner world" generated from previous experience, allowing it to try out different possible futures without executing actions overtly. We found that the success rate in terms of reaching the specified goal was higher during internal simulation, compared to overt action. These results are linked to a reduction in prediction errors generated during covert action. Despite the fact that the model is quite successful in terms of generating covert behavior toward specified goals, internal simulations display different temporal distributions compared to their overt counterparts. Links to human cognition and specifically mental imagery are discussed.

Keywords: embodied cognition, imagination, internal simulation, learning from demonstration, representation, simulation theory, predictive sequence learning, prospection

1. INTRODUCTION

Cognitive science has traditionally equated cognition with the processing of symbolic internal representations of an external world [e.g., Pylyshyn (1984), Fodor and Pylyshyn (1988), Newell (1990), and Anderson (1996)]. While clearly humans experience some kind of "inner world," i.e., the ability to imagine their environment and their own interactions with it embodied/situated theories of cognition [e.g., Varela et al. (1991), Clancey (1997), Clark (1997), and Lakoff and Johnson (1999)] have questioned the traditional view of symbolic mental representations. In artificial intelligence research, in particular, some have argued for the need of "symbol grounding" (Harnad, 1990), i.e., the grounding of amodal symbolic representations in non-symbolic iconic and categorical representations that allow to connect senses to symbols, while others have argued that the "physical grounding" of "embodied" and "situated" robots simply makes representation unnecessary [e.g., Brooks (1991)]. In this context, alternative accounts of cognition as based on different types of mental simulation or emulation have gained substantial interest [e.g., Barsalou (1999), Hesslow (2002), Grush (2004), Gallese (2005), and Svensson (2013)]. According to these theories, the "inner world" and the human capacity for imagination are based on internally simulated action and perception, i.e., the brain's ability to (re- or pre-) activate itself as if it was in actual sensorimotor interaction with the external world.

While there have been many advances in providing robots with some kind of inner world, the inner worlds of robots have traditionally been based on predefined ontologies and still lack in complexity and flexibility compared to the inner worlds of humans. In this paper, we describe a learning mechanism that enables a simulated robot to mentally imagine moving around in an apartment environment. The robot does not only repeat previously experienced routes but also shows a kind of organic compositionality (Tani et al., 2008), allowing it to reenact – and recombine – parts of previous sensory–motor interactions in novel ways. The basic mechanism underlying this is grounded in simulation theories, in particular, the type of mechanisms suggested by Hesslow (2002) and consists of learning associations between sensor and motor events.

In the present work, we combine previous efforts on internal simulation (Stening et al., 2005; Ziemke et al., 2005; Svensson, 2013) with those on *Learning from Demonstration (LFD)* (Billing, 2012) into a model that can learn from human demonstrations and reenact the demonstrated behavior both overtly and covertly. We here evaluate several aspects of such covert action: (1) can the robot produce internal simulations similar to a previously executed overt behavior; (2) to what extent can the system produce internal simulations of new behavior, that is, reenact and recombine previously experienced episodes into a novel path through the environment; and (3) how can such internal simulations in a robot be compared with simulation theories of human cognition?

The rest of this paper is organized as follows. Simulation theories of cognition are introduced in Section 2, and a problem statement of recombining previous experiences into novel simulations is presented in Section 3. The modeling technique, based on *Predictive Sequence Learning (PSL)*, is presented in Section 4, followed by a description of the experimental setup in Section 5.

Our hypotheses are made explicit in Section 6. Results are presented in Section 7, and the paper is concluded with a discussion in Section 8.

2. SIMULATION THEORIES

In simple terms, simulation theories explain cognition as simulated actions and perceptions. The term simulation, as used in this paper, is more specifically related to the following two notions:

- *Reactivation*: Simulation is the reactivation of various brain areas, especially areas along the sensory and motor hierarchies, and
- *Prediction*: The covert or simulated actions can directly evoke sensory activity that corresponds to the activity of the sensory organs that has previously occurred following action execution in that context.

Ideas relating to reactivation and prediction are not new but have received renewed attention in the last couple of decades [e.g., Damasio (1994), Barsalou (1999), Möller (1999), Hesslow (2002), Gallese (2003), and Grush (2004)]. The idea of reactivation can be found dating (at least) as far back as the British empiricists and associationists (Hesslow, 2002).

> Alexander Bain suggested that thinking is essentially a covert or weak form of behavior that does not activate the body and is therefore invisible to an external observer [...]. Thinking, he suggested, is restrained speaking or acting (Hesslow, 2002, p. 242).

The idea of restrained actions was also popular among some of the behaviorists, perhaps most prominently Watson, who viewed cognition or thinking as motor habits in the larynx (Watson, 1913, p. 84), cited in Hickok (2009). While the idea of reactivation in these early theories was rather underspecified and susceptible to criticism (for example, the finding that paralysis induced to the muscles by curare did not have any observable effects on thinking) (Smith et al., 1947; Hesslow, 2002), modern theories of simulation and reactivation [e.g., Barsalou (1999), Hesslow (2002), and Grush (2004)] further specify the nature of the reactivations (i.e., simulated actions and perceptions) based on both behavioral studies using elaborate experimental setups and a large body of neuroscientific evidence, e.g., Jeannerod (2001). We do not elaborate on the empirical evidence cited in support for simulation theories in this paper, but reviews can be found in Colder (2011), Hesslow (2012), and Svensson (2013).

Given that simulation theories have been put forward to explain many different cognitive phenomena and span a wide range of disciplines, such as linguistics [e.g., Zwaan (2003)], neuroscience (Colder, 2011), and psychology [e.g., Barsalou (1999)], they are not entirely coherent in their particular details of implementation and hypotheses about the underlying mechanisms. They also differ with regard to their view of knowledge and the relation between the cognitive agent and its environment (Svensson, 2007, 2013). However, to some extent, they share a commitment to the reactivation hypothesis and/or the prediction hypothesis. The following subsections provide a summary of three of the perhaps most commonly cited simulation theories: Hesslow's

simulation hypothesis (Hesslow, 2002), Grush's emulation theory of representation (Grush, 2004), and Barsalou's notions of perceptual symbol systems and situated conceptualizations (Barsalou, 1999, 2005).

2.1. Simulation Hypothesis

Hesslow (2002, 2012) argued that his simulation hypothesis rests on three basic assumptions:

- *Simulation of actions:* We can activate motor structures of the brain in a way that resembles activity during normal action but does not cause any overt movement.
- *Simulation of perception:* Imagining perceiving something is essentially the same as actually perceiving it, only the perceptual activity is generated by the brain itself rather than by external stimuli.
- *Anticipation:* There exist associative mechanisms that enable both behavioral and perceptual activity to elicit perceptual activity in the sensory areas of the brain (Hesslow, 2002, p. 242). Most importantly, a simulated action can elicit perceptual activity that resembles the activity *that would have occurred* if the action had actually been performed.

A central claim of the simulation hypothesis is that it is not necessary to posit some part of the brain or some autonomous agent self-performing the simulation, but the anticipation mechanism will ensure that most actions are accompanied by probable perceptual consequences (Hesslow, 2002). That means there is no need to posit an independent agent (i.e., homunculus) that evaluates the simulation; rather, the (simulated) sensory events will elicit previously learned affective consequences that guide future behavior by rewarding or punishing simulated actions. The mechanisms that ensure that the simulations are established are likely to be realized by neural mechanisms located in many different areas of the brain, rather than there being a single neural mechanism for anticipation (Svensson et al., 2009b; Svensson, 2013).

2.2. Emulation Theory of Representation

Grush (2004) proposed a general theory of representation based on the control-theoretic concept of emulation or forward modeling. The concept of a forward model is well known in motor

control and has, in that context, also been linked to seemingly more mental abilities such as forming mental images of actions [e.g., Wolpert et al. (1995)].

Generally, a forward model (ϕ) takes the current state (x_t) of the system and a control signal (u_t), and estimates the consequences of that control signal in terms of a new state of the system ($\hat{x}_{t+1}$), at some future time $t + 1$:

$$\hat{x}_{t+1} = \phi(x_t, u_t) \tag{1}$$

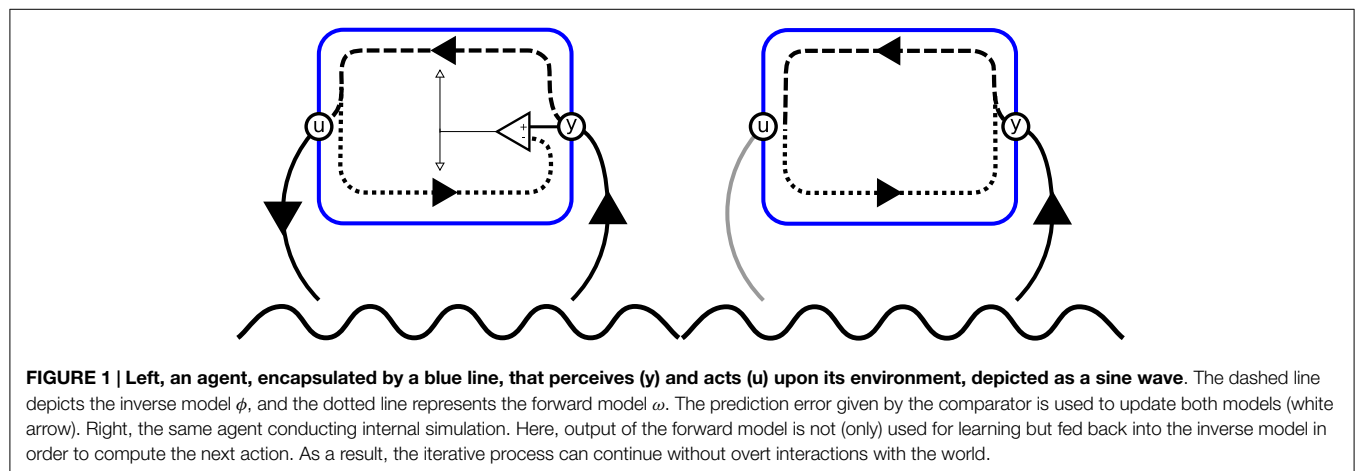The forward model acts in combination with the controller, or inverse model π:

$$\hat{u}_t = \pi(x_t) \tag{2}$$

An illustration of an agent implementing forward and inverse models along the principles put forward by Hesslow (2002) and Grush (2004) is presented in **Figure 1**. We should, however, note that the forward and inverse models are here depicted as associations between perceptions and actions, not functions of the system's state as defined in equations (1) and (2). Grush (2004) uses a Kalman filter to compute the system state based on perceptual information. However, Hesslow (2002) takes an associationist's perspective and argues for an implicit state representation.

While, as already noted above, forward and inverse models of the motor system have been linked to mental imagery, the general idea of simulation theories is that simulations are not restricted to only the immediate control of the body. As can be seen in, e.g., Barsalou (1999)'s theory of perceptual symbol systems and situated conceptualizations, simulations can also include more distal and distant aspects of embodied interaction (Svensson et al., 2009b).

2.3. Perceptual Symbol Systems

Barsalou (1999, 2005) proposed a theory of perceptual symbol systems, consisting of three parts: (1) perceptual symbols, i.e., the reenactment of modality-specific states; (2) simulators and associated simulations; and (3) situated conceptualizations. For readability, we only focus on the latter (Barsalou, 1999) for a description of parts 1 and 2. According to Barsalou (2005), our ability to categorize and conceptualize the world depends on a particular type of simulation, which he terms situated conceptualizations,

FIGURE 1 | Left, an agent, encapsulated by a blue line, that perceives (y) and acts (u) upon its environment, depicted as a sine wave. The dashed line depicts the inverse model ϕ, and the dotted line represents the forward model ω. The prediction error given by the comparator is used to update both models (white arrow). Right, the same agent conducting internal simulation. Here, output of the forward model is not (only) used for learning but fed back into the inverse model in order to compute the next action. As a result, the iterative process can continue without overt interactions with the world.

in which the conceptualizer is placed directly in the respective situations, creating the experience of *being there* . . . (Barsalou, 2005, p. 627). Barsalou illustrated this as follows:

> Consider a situated conceptualization for interacting with a purring house cat. This conceptualization is likely to simulate how the cat might appear perceptually. When cats are purring, their bodies take particular shapes, they execute certain actions, and they make distinctive sounds. All these perceptual aspects can be represented as modal simulations in the situated conceptualization. Rather than amodal redescriptions representing these perceptions, simulations represent them in the relevant modality-specific systems (Barsalou, 2005, p. 626–627).

In such a situation, simulated perceptions and actions/emulations are connected into simulated chains of embodied interactions that involve bodily aspects as well as physical and social aspects of the environment and enable our conceptual understanding of the situation.

3. PROBLEM STATEMENT

One of the key premises of simulation theories is that the capacity to reenact previous experiences covertly allows the agent to get away from the *here and now* and generate the experience of a novel sequence of events. This allows the agent to "try out" different scenarios without the effort and possible dangers of actually executing them.

As an example, *close your eyes and imagine yourself in your home. You wake up, get out of the bed, go through the bedroom door, continuing the shortest way out of the building* – it is quite easy to imagine, even if you have never taken exactly this path before. You combine previous experiences from your home into something new.

An abstraction of the scenario described above is given in **Figure 2**. With the knowledge of getting from A to B and from C to D, the internal simulation exploits that intersection to postulate a possible way of getting from A to D or C to B. An agent standing at A can of course follow the known path from A toward B and look for opportunities to reach D, but with many behaviors to choose from the goal quickly becomes unreachable. Internal simulations drastically reduce the effort of trying out different combinations of known behaviors in the real world.

Despite the fact that simulation theories constitute common sources of inspiration for roboticists, the basic scenario described

FIGURE 2 | Abstract depiction of two paths, from A to B and from C to D. The two paths, depicted as dashed lines, share some experiences, indicated with a dotted circle, allowing the agent to combine previously experienced episodes into new behavior, e.g., going from A to D. See text for details.

above has to our knowledge never been computationally analyzed using internal simulation. Over the last two decades, computational models, such as HAMMER (Demiris and Hayes, 2002; Demiris and Khadhouri, 2006; Demiris et al., 2014), MOSAIC (Wolpert and Kawato, 1998; Haruno et al., 2003), MTRNN (Yamashita and Tani, 2008), and MSTNN (Jung et al., 2015), have received significant attention. These models all rely on prediction and to some extent a pairing of forward and inverse models, and some have been used for internal simulation. But none have, to our knowledge, been used for generating novel, goal-directed behavior using internal simulation.

Tani (1996) presents an early approach to *model-based learning* using a recurrent neural network that generates an internal simulation of future sensory input. Another example of goal-directed planning using internal simulations was presented by Baldassarre (2003). While both these models produce simulations of goal-directed behavior, they do not learn from human demonstrations.

Pezzulo et al. (2013) argued that there is a need for grounded theories, including simulation theories, to develop better process models of "how grounded phenomena originate during development and learning and how they are expressed in online processing" and that an important challenge is explaining "how abstract concepts and symbolic capabilities can be constructed from grounded categorical representations, situated simulations, and embodied processes." The grounded cognition approach and Barsalou's situated conceptualization (Barsalou, 1999, 2005; Barsalou et al., 2003) suggest that abstract thinking and other cognitive phenomena are based on and consist of processing involving bodily (e.g., proprioceptive, interoceptive, and emotional) states as well as the physical and social environment. Therefore, computational models of simulation theories need to better reflect the bodies and environments that humans inhabit to develop richer concepts that can be used to think about the world and oneself.

Here, we directly address the latter part of this question by evaluating to what extent the robot model, based on the principles of simulation theories, can internalize the environment and generate new behavior in the form of internal simulations. Specifically, we are interested in to what degree the robot model can produce goal-directed covert action and to what degree the internal simulation replicates the sensory–motor interactions of the overt behavior.

3.1. Computational Models of Internal Simulation

A possible approach of implementing simulations as proposed by Hesslow (2002) in computational agents can be found already in the work of Rumelhart et al. (1986) on artificial neural networks. They suggested that it is possible to "run a mental simulation" by having one network that produces actions based on sensory input and another that predicts how those actions change the world. By replacing the actual inputs with the predicted inputs, the networks would be able to simulate future events (Rumelhart et al., 1986, p. 41–42). The assumption is that if the predicted sensory input is similar enough to the actual sensory input, the agent can be made to operate covertly where the predicted sensory input is used instead of the actual sensory input. Two early implementations of simulation-like mechanisms of this kind were the "Connectionist
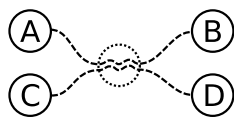
Navigational Map (CNM)" of Chrisley (1990) and the simulated robot (arm) "Murphy" by Mel (1991). The theoretical motivation behind the connectionist navigational map was the transition from non-conceptual to conceptual knowledge [see, e.g., Barsalou (1999)], and later models have focused on, for example, allocentric spatial knowledge (Hiraki et al., 1998), learning the spatial layout of a maze-like environment [e.g., Jirenhed et al. (2001) and Hoffmann and Möller (2004)], obstacle avoidance (Gross et al., 1999), and robot dreams (Svensson et al., 2013).

These experiments have shown that the computational agents can learn to produce internal simulations that guide behavior in the absence of sensory input. However, it is not a trivial task to construct such simulations. For example, one can easily imagine that if predictions start to diverge from the actual sensory states, simulations will drift and become increasingly imprecise. Jirenhed et al. (2001) found that some behaviors caused states in which predictions could not be learnt which hindered successful internal simulations to develop, and Hoffmann and Möller (2004) identified an accumulation of error as the chains of predictions increase in length, which could restrict the ability to create longer sequences of simulations. On the other hand, in Baldassarre (2003)'s model, noisy predictions did not accumulate for each time step. Others have suggested that the states in internal simulations should not be judged by their correspondence to real sensory input, rather the important aspect is that the internal simulations constructed by the robot support successful behavior (Gigliotta et al., 2010). For example, Ziemke et al. (2005) demonstrated a simple internal simulation in a Khepera robot [K-Team (2007)], using a feed forward neural network. The network generated both actions and expected percepts, allowing the robot to reenact earlier sensory–motor experiences and move blindfolded through its environment. Despite the fact that the actions produced during internal simulation produced roughly the same coherent behavior, the internally generated sensor percepts used for blind navigation were actually quite different from the previously experienced real sensory inputs.

The current models of mechanisms based on simulation theories have shown the viability of creating internal models out of simple sensory–motor associations, but the environments have often been of very low complexity. For example, the environments consist of very simple mazes without obstacles/objects [e.g., Jirenhed et al. (2001) and Ziemke et al. (2005)] or with only a few simple block shaped obstacles/objects [e.g., Tani (1996) and Gross et al. (1999)].

Two computational models that have been used in more complex settings are HAMMER (Demiris and Simmons, 2006; Demiris et al., 2014) and the recurrent neural network models by Tani and colleagues (Tani et al., 2008; Yamashita and Tani, 2008; Jung et al., 2015).

Already in an early presentation of the model (Demiris and Hayes, 2002), HAMMER was used for internal simulation in the context of imitation learning. Framed as *active imitation*, Demiris and Hayes (2002) present a system producing a set of parallel internal simulations. The output of each simulation, generated by paired forward and inverse models, is compared with the demonstrator's observed state, and the error is used to assign a confidence level to each simulation. Since each pair of forward and inverse models is trained to implement a specific behavior,

the method can be used for imitation. Together with the MOSAIC model (Haruno et al., 2003), HAMMER constituted significant inspiration for our own previous work on *behavior recognition* (Billing et al., 2010). For further positioning of the HAMMER architecture in relation to other cognitive architectures, please refer to Vernon (2014).

Another topic of interest has been the formation of concepts and abstractions of the sensorimotor flow in internal simulations. For example, Stening et al. (2005) developed a two-level architecture in which the higher level was able to form internal simulations of the "rough" structure of the environment, based on simple categories, such as "corner" or "corridor," developed through unsupervised learning at the lower level. Another example is Tani et al. (2008), who investigated the development of flexible behavior primitives achieving a kind of "organic compositionality" in a humanoid robot, and the robot was also able to reactivate the primitives internally in a mental simulation. These types of behavior primitives could be seen as potential buildings blocks of the situated conceptualizations suggested by Barsalou (2005).

4. PREDICTIVE SEQUENCE LEARNING

As discussed in Section 3, we are interested in investigating how prior experiences can be combined into new, goal-directed behavior during both overt and covert actions. The proposed system is based on our earlier work on Learning from Demonstration (LFD) (Billing and Hellström, 2010) and the PSL algorithm [e.g., Billing et al. (2011a,b, 2015)]. PSL resembles many of the principles put forward in Sections 2–3 and implements forward and inverse models as a joint sensory–motor mapping:

$$\hat{e}_{t+1} = f(\eta_t) \tag{3}$$

where $e_t = (u_{t-1}, y_t)$ represents a *sensory–motor event* at time t, comprising perceptions y_t and actions u_{t-1}. $\hat{e}_{t+1}$ is the predicted event estimate. A sequence of events constitutes the sensory–motor event history η:

$$\eta_t = (e_1, e_2, \ldots, e_t) \tag{4}$$

PSL constitutes a minimalist approach to prediction and control, compared to, e.g., HAMMER (Demiris and Hayes, 2002; Demiris and Khadhouri, 2006) and CTRNN (Yamashita and Tani, 2008), as discussed in Section 2. Both HAMMER and CTRNN have been evaluated in LFD settings and could be expected to produce more accurate prediction and control, especially with high-dimensional and noisy data. However, for the present evaluation, PSL was chosen since it, in contrast to the HAMMER architecture, represents a fully defined algorithm, leaving less room for platform specific interpretations. It also takes an associationist approach to learning, implementing a direct perception – action mapping closely resembling Hesslow's simulation hypothesis (Hesslow, 2012) as depicted in **Figure 1** [compared with the use of a state estimate, equations (1) and (2)]. In this respect, PSL is model free (Billing et al., 2011a) and, in the form used here, comprises only two parameters: *membership function size* (τ) and the *precision constant* $\hat{\alpha}$. Both parameters are described in detail below.

With a closer connection to biology, CTRNN represents a theoretically interesting alternative but requires much larger training times and could therefore pose practical problems for conducting the kind of evaluations presented here.

In the language of control theory, we would define a transfer function describing the relation between the system's inputs and outputs. An estimator, such as maximum likelihood, would then be used to estimate the parameters for the model. For an introduction to this perspective in robotics, see, e.g., Siegwart and Nourbakhsh (2004). Here, we take a different approach and formulate f [equation (3)] as a set of fuzzy rules, referred to as *hypotheses* (h), describing temporal dependencies between a sensory–motor event e_{t+1} and a sequence of past events ($e_{t-|h|+1}$, $e_{t-|h|+2}, \ldots, e_t$), defined up until current time t:

$$h : \left(\Upsilon_{\tau+1} \text{ is } E_1^h \wedge \Upsilon_{\tau+2} \text{ is } E_2^h \wedge \ldots \wedge \Upsilon_t \text{ is } E_{|h|}^h \right) \overset{C}{\Rightarrow} \Upsilon_{t+1} \text{ is } \bar{E}^h. \quad (5)$$

where Υ_i is the event variable, and $E^h(e)$ is a fuzzy membership function returning a membership value for a specific e_t. The right-hand side $\bar{E}^h$ is a membership function comprising expected events at time $t + 1$. $|h|$ denotes the length of h, i.e., the number of left-hand-side conditions of the rule. τ equals $t - |h|$. C represents the confidence of h within a specific context, described in the following section. Both E and $\bar{E}$ are implemented as standard cone membership functions with base width ε [e.g., Klir and Yuan (1995)].

A set of hypotheses is used to compute f [equation (3)], producing a prediction $\hat{e}_{t+1}$ given a sequence of past sensory–motor events η. The process of matching hypotheses to data is described in Section 1, and the use of PSL as forward and inverse models during overt and covert actions is described in Section 3.

As hypotheses represent weighted associations between a sequence of sensory–motor events, PSL can be viewed as a variable-order Markov model. Generated hypotheses are initially associating a single e_t with $\hat{e}_{t+1}$, implementing a first-order association. In cases where e_t does not show Markov property, an additional hypothesis is generated: $(e_{t-1}, e_t) \Rightarrow \hat{e}_{t+1}$, implementing a second-order association. Seen as a directed graph between sensory–motor events, PSL implements several aspects of a joint procedural–episodic memory (Vernon et al., 2015). For a detailed description of the learning process of PSL, please refer to Billing et al. (2015).

PSL is not expected to produce a better estimate in terms of prediction error than what can be gained using control theory approaches, but allows estimated mapping functions, defined as sets of hypotheses, to be recombined in order to produce novel behavior. This property is not as easily achieved using a control-theoretic approach.

4.1. Matching Hypotheses

Given a sequence of sensory–motor events, $\eta = (e_1, e_2, \ldots, e_t)$, a match $\alpha_t(h)$ is given by

$$\alpha_t(h) = \bigwedge_{i=0}^{|h|-1} E_i^h(e_{t-i}) \quad (6)$$

where $\wedge$ is implemented as a *min*-function.

Hypotheses are grouped into fuzzy sets C whose membership value $C(h)$ describes the *confidence* of h at time t:

$$C(h) = \frac{\sum_{k=t_h}^{t} \alpha_k(h) \bar{E}^h(e_{k+1})}{\sum_{k=t_h}^{t} \alpha_k(h)} \quad (7)$$

where t_h is the creation time of h. Each set C represents a *context* and can be used to implement a specific behavior or part of a behavior. The *responsibility signal* $\lambda_t(C)$ is used to control which contexts are active at a specific time. The combined confidence value $\tilde{C}_t(h)$, for hypothesis h, is a weighted average over all C:

$$\tilde{C}_t(h) = \frac{\sum_{C} C(h) \lambda_t(C)}{\sum_{C} \lambda_t(C)} \quad (8)$$

where $\tilde{C}_t$ is a single fuzzy set representing the combination of all active contexts at time t. Hypotheses contribute to a prediction in proportion to their membership in $\tilde{C}$, their length, and match $\alpha_t(h)$. The aggregated prediction $\hat{E}(e_{t+1})$ is computed using the Larsen method [e.g., Fullér (2000)]:

$$\hat{E}(e_{t+1}) = \bigvee_{h} \bar{E}_h(e_{t+1}) \tilde{C}(h) |h| \alpha_t(h)^2 \quad (9)$$

During learning, new hypotheses are created when $\hat{E}(e_{t+1}) < \hat{\alpha}$, that is, when the observed sensory–motor event e_{t+1} in the training data does not match the prediction. The precision constant $\hat{\alpha} \in [0, 1]$ is, in fuzzy-logic terms, an α-cut, i.e., $\hat{\alpha}$ specifies a threshold for prediction precision, where a high value results in highly precise predictions and a large number of hypotheses, while a small $\hat{\alpha}$ renders less hypotheses, and less precise predictions [see Billing et al. (2011b) for details].

While the PSL algorithm used here is identical to earlier work, a different defuzzification method is used. Billing et al. (2015) employed a *center of max* defuzzification method, while we here use a probabilistic approach. $\hat{E}$ is treated as a probability distribution and converted to crisp values by randomly selecting a predicted sensory–motor event $\hat{e} \in \hat{E}$ in proportion to their membership in $\hat{E}$.

The PSL mapping function [equation (3)] can now be redefined as context-dependent forward and inverse models:

$$\hat{y}_{t+1} = \phi(\eta_t, \lambda_t) \quad (10)$$

$$\hat{u}_t = \pi(\eta_t, \lambda_t) \quad (11)$$

4.2. Illustrative Example

Figure 3 presents an example of PSL applied to a simplified robot scenario. Consider a robot placed in an environment depicted as a gray area in the figure. A demonstration (η), comprising 8 sensory motor events, shows how to get from its current location to the goal (G). Using η as training data, PSL generates a knowledge base comprising 9 hypotheses, under context C.

When placed in the test environment, PSL is used as a controller to generate a sequence of actions from start to goal. Note that the

starting location is not identical and that the test environment is slightly different from the one used for training. The output produced by PSL is presented in **Figure 3** as *f*, aligned with the event sequence η_r observed while executing selected actions. At time step 1, PSL bases its prediction on a single (blue) sensory–motor event which, according to the knowledge base, can have three possible outcomes (h1, h2, and h5). PSL selects among these in proportion to the confidence levels, represented by the number above hypotheses' arrows in the figure. The action associated with the selected sensory–motor event is executed and the robot approaches the corner at time step 2. Predictions are again based on η_r, h3 is selected, and a right turn is issued.

While continuing this thought example, PSL will produce correct predictions at t = 3 and 4, incorrect prediction at t = 5 and 6, and finally catching up with correct predictions at t = 7 and 8. Errors are only perceptual and appropriate actions are executed also in these cases, allowing the robot to stay on path also when there are differences between predicted and observed sensory–motor events.
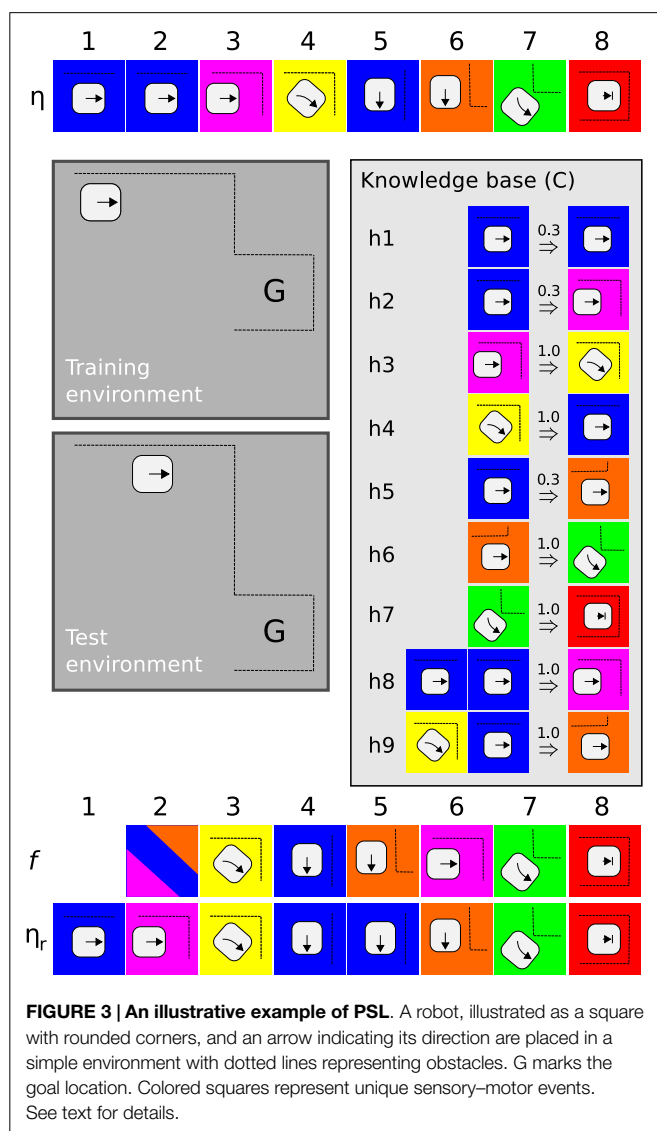

FIGURE 3 | An illustrative example of PSL. A robot, illustrated as a square with rounded corners, and an arrow indicating its direction are placed in a simple environment with dotted lines representing obstacles. G marks the goal location. Colored squares represent unique sensory–motor events. See text for details.

In a realistic scenario, there will never be an exact match between hypotheses in the knowledge base and observed perceptions and actions. Matching perceived data to sensory–motor events is controlled by the membership function [equation (5)]. A wide membership, with a large τ, allows many hypotheses to be selected, increasing the robot's ability to act also in relatively novel situations. However, a value of τ that is too large reduces precision as a larger number of hypotheses match observed data, increasing the risk that inappropriate actions are selected even in well-known environments. A balance between the two allows certain variability in the environment, while still producing stable behavior. This balance can hence be seen as a type of exploration – exploitation trade-off present in many machine learning approaches.

4.3. Overt and Covert Actions

Hypotheses generated by PSL are used in two modes: (1) as a robot controller (overt action) and (2) for internal simulation (covert action). In Mode 1, the forward model is ignored, and π_{psl} [equation (11)] is directly used as a controller for the robot. All sensory–motor events e_t comprises perceptions y_t from the robot's sensors and actions $u_t = \pi_{psl}(\eta_t, \lambda_t)$. This process resembles **Figure 1** (left) with the distinction that learning is only active when the robot is teleoperated by a human teacher.

Mode 2 resembles the right part of **Figure 1**. π_{psl} is here paired with ϕ_{psl} (Eq. 10) to create a reentrant system. Here, only $e_1 = (\emptyset, y_1)$ is taken from the robot's sensors. All events $\hat{e}_{t>1}$ are generated by

$$\hat{e}_{t+1} = \left(\pi_{psl}\left(\eta_t, \lambda_t \right), \phi_{psl}\left(\eta_t, \lambda_t \right) \right) \quad (12)$$

$$\eta_t = (e_1, \hat{e}_2, \ldots, \hat{e}_t) \quad (13)$$

As a result, the internal simulation is only based on e_1, the demonstrations used to train PSL, and the responsibility signal λ_t. While λ is in general time varying and can be computed dynamically, using a method for behavior recognition (Billing et al., 2011b, 2015), it was here used as a constant goal signal.

For analytic purposes, we also need to define the *prediction error* $\delta_t(C)$:

$$\delta_t(C) = 1 - \hat{E}(C, e_t) \quad (14)$$

representing the error for context C at time t. $\hat{E}(C, e_t)$ denotes the context specific aggregate [equation (9)]. As mentioned above, $e_{t>1}$ is not available during covert action. In this case, we consider $e_t \approx \hat{e}_t$, allowing computations of prediction errors also during internal simulation.

Based on our measure of prediction error, the *confidence* $\gamma_t(C)$ in context C at time t is given by

$$\gamma_t(C) = \frac{\gamma_{t-1}(C) \exp\left(\frac{(E(C, e_t) - 1)^2}{2\lambda^2} \right)}{\sum_{i=1}^{N} \left[\gamma_{t-1}(C_i) \exp\left(\frac{(E(C_i, e_t) - 1)^2}{2\lambda^2} \right) \right]} \quad (15)$$

This definition of confidence has its roots in the MOSAIC architecture (Haruno et al., 2003) and has previously been used to compute the responsibility signal $\lambda_t = \gamma_t$ online (Billing et al., 2015).

5. EXPERIMENTAL SETUP

To evaluate to what degree the model is able to produce goal-directed behavior during both overt and covert actions, nine test cases were evaluated. A simulated Kompaï robot (Robosoft, 2011) placed in an apartment environment (**Figure 4**) was selected as a test platform. Microsoft Robotics Developer Studio (MRDS) was used for robot simulations. The apartment environment is a standard example environment, freely available from Microsoft (2015). PSL[1] and related software was implemented using Java™. Motivation and hypotheses follow in Section 6.

The Kompaï robot was equipped with a 270° laser scanner and controlled by setting linear and angular speeds converted to motor torques by the low-level controller. The 271 laser scans were converted into a 20-dimensional vector where each element represents the mean distance within a 13.5° segment of the laser data. In total, each sensory–motor event e comprised 20 sensor dimensions from the laser data and two motor dimensions (linear and angular speeds). All data were sampled over 20 Hz, and as a result, each sensory–motor event had a temporal extension of 50 ms.

A similar setup was used in previous evaluations of PSL (Billing et al., 2011b, 2015). While the setup is still far from humanoid sensor and motor complexity, the Kompaï robot is designed to act in human environments and does represent a significant increase in environmental and perceptual complexity compared to previous work using simulated Khepera robots (Jirenhed et al., 2001; Stening et al., 2005; Ziemke et al., 2005; Svensson et al., 2009a).

---

[1]The Java™ implementation of Predictive Sequence Learning and related libraries are freely available as a software repository at https://bitbucket.org/interactionlab/psl, licensed under GNU GPL3.

Three *behaviors* with different start and goal locations were demonstrated by remote controlling the robot using a joy pad:

- *ToKitchen*: From the storage room (Area 1) to the kitchen (Area 2);
- *ToTV*: From the bed (Area 3) to the TV (Area 4);
- *GoOut*: From the bathroom (Area 5) to the elevator (Area 6).

Each behavior was demonstrated four times, producing a total of 12 demonstrations. During demonstration, sensor readings and executed motor commands were recorded. Laser scans were given a maximum distance of 16 m and the membership function base (ε) was set to 1.6 m. $\hat{\alpha} = 0.9$ for all conditions. Parameter selection was based on previous work (Billing et al., 2015), where a similar setup was used. Preliminary demonstrations were recorded to select suitable start and stop locations and to verify the technical implementation. These demonstrations were thereafter discarded. A set of 12 demonstrations was then recorded for the final training set. These demonstrations were verified by training PSL on all demonstrations from a single behavior and letting PSL act as a controller for the robot, reproducing the demonstrated behavior. All demonstrations passed verification; hence, no recordings were discarded.

For the following test cases, all 12 demonstrations were used as training data for PSL. During training, each behavior was given a unique context [equation (7)], allowing a top-down responsibility signal [equation (8)] to bias selection of hypotheses when PSL was used as a controller. As a result, the selection of specific context can be said to indicate a goal in the form of the target location of the corresponding behavior. PSL was trained on all 12
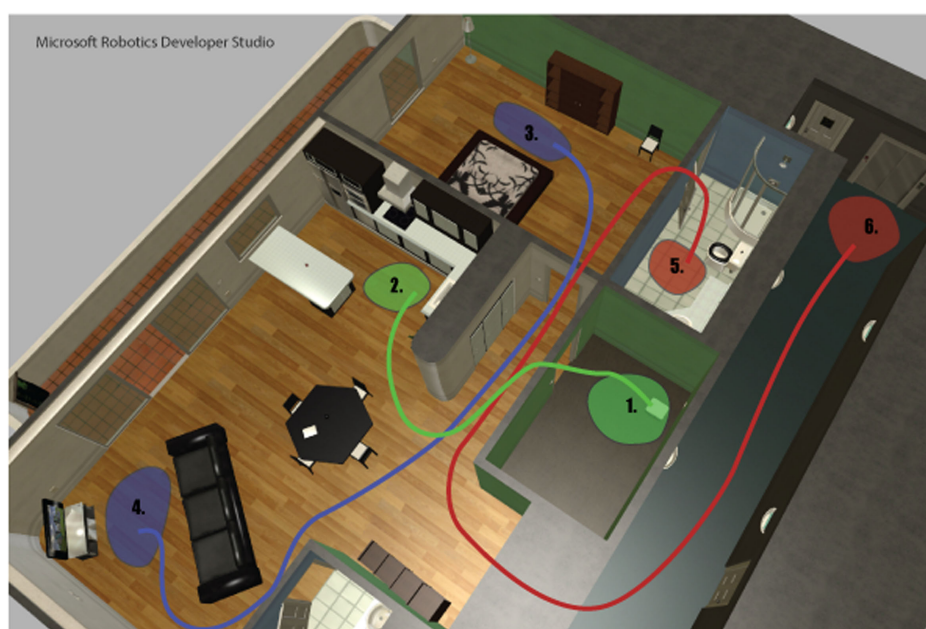

FIGURE 4 | Simulated apartment environment. Green, blue, and red lines indicate approximate paths from start (1, 3, and 5) to goal (2, 4, and 6) locations, for the three demonstrated behaviors, *ToKitchen*, *ToTV*, and *GoOut*. See text for details.

demonstrations for eight epochs each. One epoch is here defined as a single presentation of all 12 demonstrations[2] in random order.

In each test phase, the robot was placed on one of the three starting locations (Areas 1, 3, and 5, **Figure 4**) and executed with a top-down responsibility signal [$\lambda(C)$] selecting one of the three goal locations presented during the demonstration phase. For selected context, $\lambda(C) = 1.0$. The responsibility signal for other contexts was set to $\lambda(C) = 0.1$, allowing hypotheses from these contexts to influence prediction, but down-prioritized in competition with hypotheses trained under the selected context. See Section 4 for a detailed mathematical formulation.

All combinations of the three starting positions and three goals (contexts) were tested, constituting a total of nine conditions: storage to kitchen (ToKitchen), bed to kitchen, bathroom to kitchen, storage to TV, bed to TV (ToTV), bathroom to TV, storage to elevator, bed to elevator, and bathroom to elevator (GoOut). Note that ToKitchen, ToTV, and GoOut were the behaviors used in the demonstration and training phase. Each condition was executed 20 times using overt action and another 20 times using covert action (internal simulation), producing a total of 360 trials[3]. See Section 4.3 for a detailed description of the two modes of execution.

6. HYPOTHESES

Following the basic premise presented in Section 3, successful internal simulation should be able to produce realistic sensory–motor interactions of a novel path through the environment.

H1: the simulated robotic system should be able to reenact all nine conditions presented in Section 5, producing an internal simulation connecting the sensory–motor state perceived at the

---

[2]Repeated presentation of the same training data allows PSL to form stable statistical dependencies between sensory–motor events and to extend the temporal window, i.e., creating longer hypotheses, when needed. This repeated presentation of sensor data may not be completely realistic from a biological point of view but can be seen as a standard method similar to the many epochs used when training, e.g., artificial neural networks.

[3]Log files from human demonstrations and all 360 simulated trails are available for download at https://bitbucket.org/interactionlab/psl/branch/reenact

starting point, with the sensory–motor state corresponding to the goal.

In humans and animals, internal simulations happen at different temporal scales (Svensson et al., 2009b; Svensson, 2013), comprising automatic unconscious mental simulations involved in, for example, perception that occur at a very rapid time scale and often involve detailed sensor and motor states [e.g., Gross et al. (1999), Möller (1999), and Svensson et al. (2009b)]. Deliberate mental simulations, e.g., mental imagery, occur at time scales corresponding to the overt behavior (Guillot and Collet, 2005). For example, Anquetil and Jeannerod (2007) studied humans performing mental simulations of grasping actions in both first and third person perspectives. The time to complete simulated actions was found to be closely similar in the two conditions. The approach used here runs internal simulations solely on a sensory–motor level, with exactly the same speed (20 Hz) as the overt behavior.

H2: internal simulations are therefore expected to display a similar temporal extension as the corresponding overt behavior.
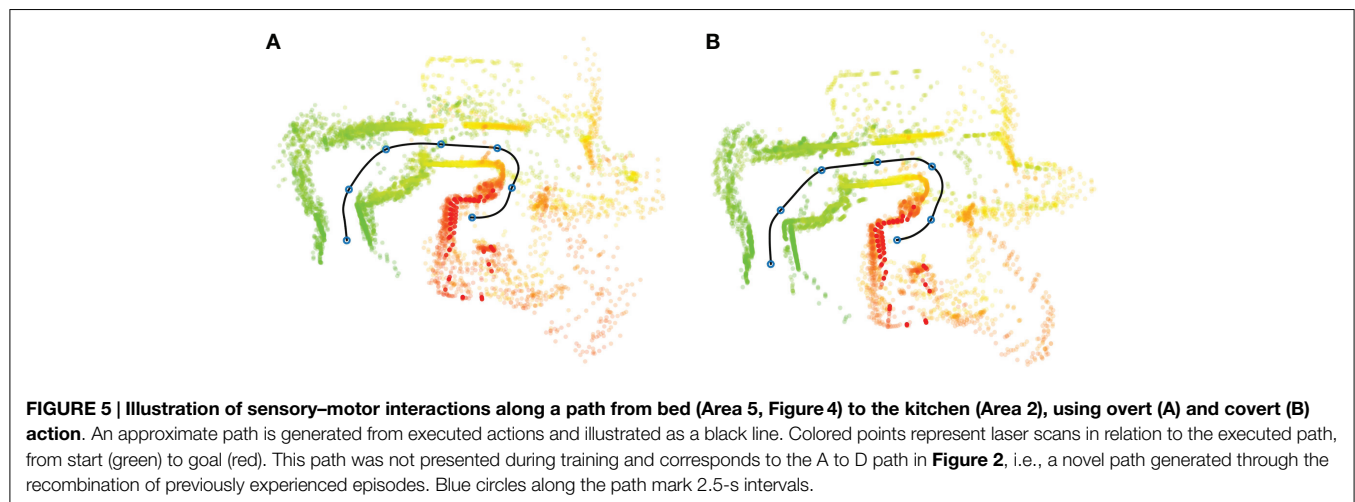
7. RESULTS

Figure 5A presents sensor perceptions (laser scans) generated during overt action, plotted in relation to the executed path from the bed (Area 5, **Figure 4**) to the kitchen (Area 2). While both bed and kitchen were present as start and goal locations in the training data, the path from bed to kitchen was not demonstrated. The robot model combines previously experienced episodes from several different demonstrations into a novel path, corresponding to the schematic illustration from A to D presented in **Figure 2**.

The path from bed to kitchen was also executed covertly (**Figure 5B**). In this case, actions were not sent to the robot controller and the path (black line) was reconstructed from the sequence of covert actions. Presented sensor percepts are not taken from the robot's sensors, but instead generated by the internal (PSL) model.

As an illustration of how information from the different demonstrations were used during internal simulation, prediction

FIGURE 5 | Illustration of sensory–motor interactions along a path from bed (Area 5, Figure 4) to the kitchen (Area 2), using overt (A) and covert (B) action. An approximate path is generated from executed actions and illustrated as a black line. Colored points represent laser scans in relation to the executed path, from start (green) to goal (red). This path was not presented during training and corresponds to the A to D path in **Figure 2**, i.e., a novel path generated through the recombination of previously experienced episodes. Blue circles along the path mark 2.5-s intervals.

errors, and confidence levels for each context (behavior) is presented in **Figure 6**. As visible in the figure, the *ToTV* context is initially generating relatively small prediction errors, leading to high confidence levels for this context (c.f. Section 3). After about 10 s, *ToTV* is starting to produce larger errors, leading to a switch in confidence to the *ToKitchen* context. This switch is the result of a strong responsibility signal $[\lambda_t(C)]$ for the *ToKitchen* context and also associated with the robot turning toward the kitchen (c.f. **Figure 5**).

Both examples presented in **Figure 5** are successful in the sense that the correct goal, indicated by the top-down signal λ, was reached. Over all nine conditions executed with overt action, the correct goal was reached in 65% of all runs. In 29% of the runs, one of the remaining two goals was reached, leaving 6% of the runs failed, in the sense that no goal was reached.

In runs with covert action, the robot did not move. In order to determine how the internal simulation terminated, the sensory–motor patterns of all 180 covert runs were plotted (as in **Figure 5B**) and the goal was visually identified. Over all nine conditions, the correct goal was reached in 75% of all runs, a different goal was reached in 7% of the runs and 18% of the runs were classified as failed.

Goal reaching frequencies for each condition, including both overt and covert runs, are presented in **Figure 7**. Examples of failed internal simulations and internal simulations reaching the wrong goal are given in **Figure 8**.

7.1. Simulation Time

In order to test hypothesis 2 (Section 6), simulation time is compared to the time of the overt behavior, presented in **Figure 9**. Some conditions, *Bed to Kitchen* and *Bathroom to TV*, display similar temporal distributions. However, seen over all conditions, the correlation between overt and covert execution times is weak, with internal simulations producing both longer (*Storage to TV*)
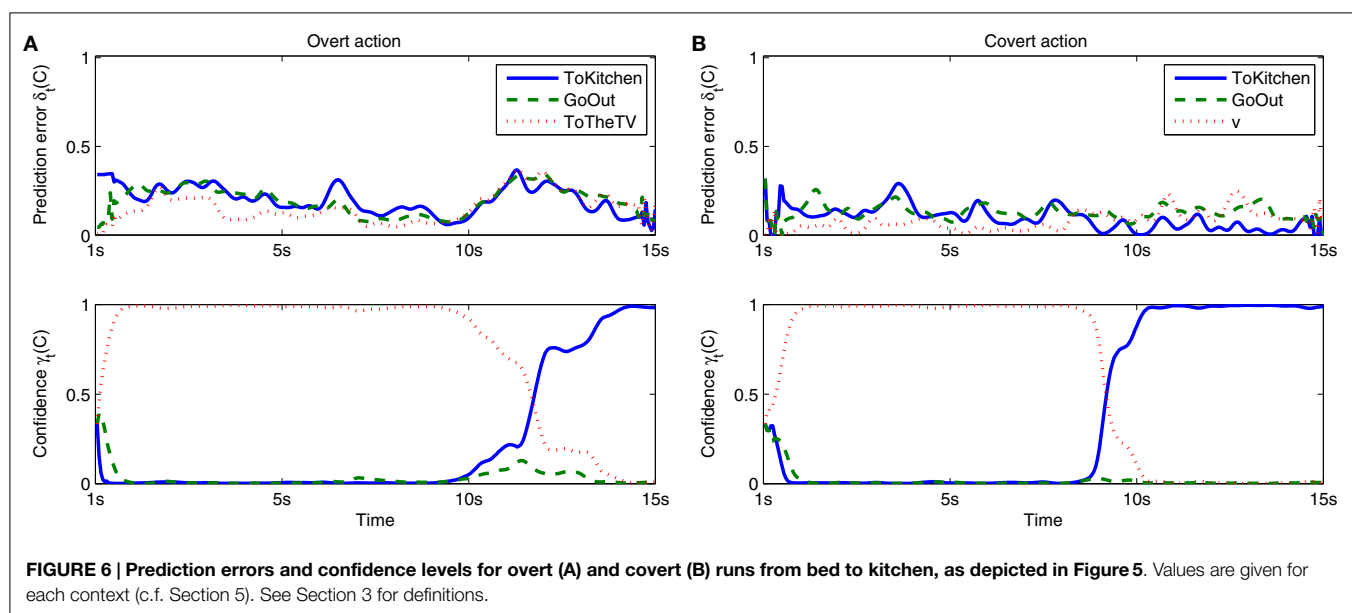
and shorter (bottom three conditions, **Figure 9**) execution times. A two-tailed *t*-test over all runs reveals a significant difference between overt and covert execution times (p < 0.005).

The strongest difference between overt and covert execution times is found in conditions *Storage to TV* and *Bed–Elevator*, with the former showing longer covert execution times, and the latter shorter times for covert runs. A deeper analysis of these conditions is presented below.

Typical runs from *Storage to TV* are displayed in **Figure 10**. The internal simulation (b) is semantically correct; it replicates all important aspects of the overt execution (a), but misrepresents the first part of the path, through the storage room depicted in green. The total execution time of the simulated path is in this case almost twice as long as its overt counterpart, 30 versus 18 s. The time of exit from the storage room is 17.5 s in the covert condition and 4.0 s in the overt case, implying that the vast majority of the temporal difference between the two conditions appears in the storage room.

This distortion of the internal simulation could not be explained by a difference in prediction error. A two tailed *t*-test showed no significant difference between prediction errors for overt and covert conditions over the relevant periods, $0 < t < 4$ s and $0 < t < 7.5$ s, respectively (p = 0.24). The observed distortion may instead be explained by a closer analysis of the PSL model. **Figures 10C,D** depicts number of matching hypotheses over time. The number of matching hypotheses is here defined as number of h for which $\alpha_t(h) > 0$ [c.f. Equation (6)]. A large number of matching hypotheses indicate larger uncertainty in the model. Both overt and covert runs display an initial period where a relatively large number of hypotheses match present sensory–motor events. Both these periods roughly correspond to the time it took to exit the storage room, 4 and 17.5 s, respectively.

In overt mode, large model uncertainty is not a problem as long as suitable actions are selected. Events are driven forward through interaction with the environment. However, in
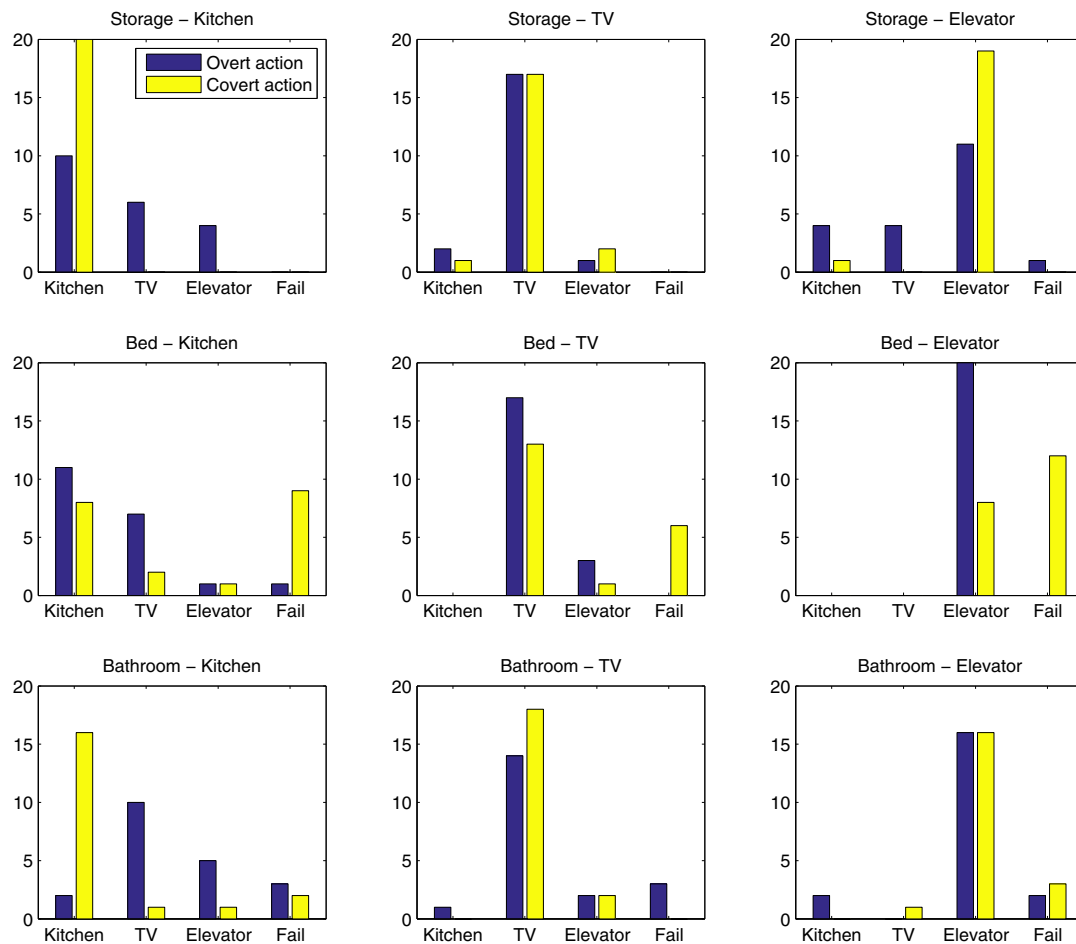

FIGURE 6 | Prediction errors and confidence levels for overt (A) and covert (B) runs from bed to kitchen, as depicted in Figure 5. Values are given for each context (c.f. Section 5). See Section 3 for definitions.

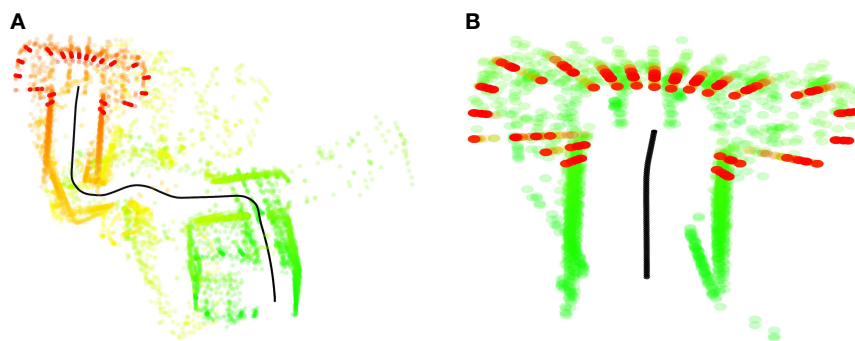**FIGURE 7 | Goal reaching frequencies for each condition**.

FIGURE 8 | (A) Internal simulation from the storage room (Area 1, **Figure 4**), here depicted in green, to the TV (Area 4), depicted in red. The simulation diverges from the path indicated by the active context (ToTV) and terminates in a sensory–motor pattern more similar to the elevator (Area 6). **(B)** Unsuccessful internal simulation from bed (Area 5, **Figure 4**) to the kitchen (Area 2). The simulation appears to mistake the bed for the entrance corridor and terminates quickly in front of the elevator.

covert mode, the PSL model must also produce suitable perceptions in order to drive the internal simulation. A larger number of matching hypotheses is likely to produce oscillating perceptions, leading to simulation distortion. This explanation is further supported by significantly larger prediction errors being generated in the storage room, compared to the rest of the executed path. This difference was observed during covert simulation ($p < 0.005$), but not for the overt condition ($p = 0.14$).
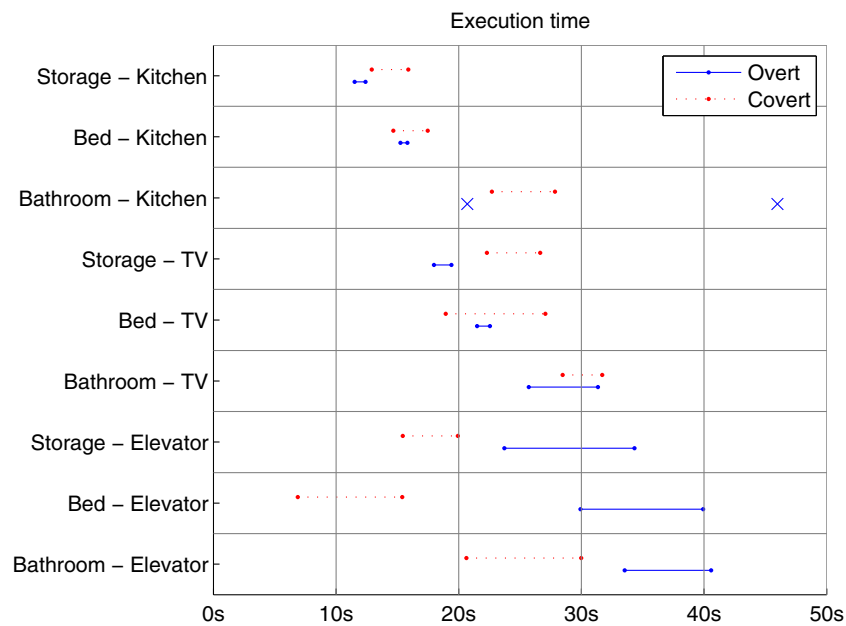
FIGURE 9 | Execution time, i.e., time to reach the goal, from all successful runs in each condition. Solid and dashed lines represent the 95% confidence interval of the means of overt and covert runs, respectively. Conditions with few successful runs are displayed as × marking execution times for individual runs, rather than a distribution.
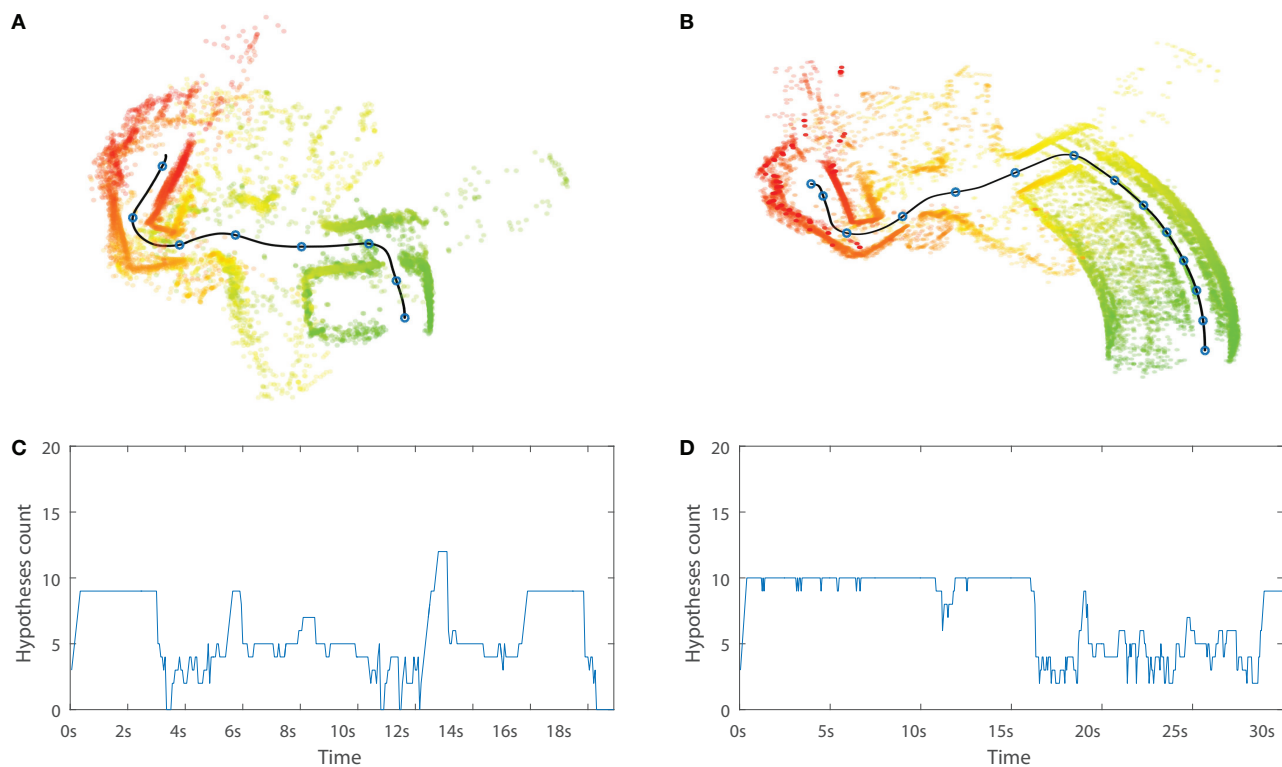

FIGURE 10 | Typical runs from condition Storage to TV. Upper plots present laser scans in relation to the executed path, generated during overt **(A)** and covert **(B)** runs from storage room (green) to TV (red). Blue circles along the path, represented by the black line, mark 2.5-s intervals. Lower plots present number of matching hypotheses over time, for overt **(C)** and covert **(D)** runs. See text for details.

A similar analysis of *Bed–Elevator* (**Figure 11**) reveals the opposite effect. In this case, the covert execution reproduces both start and goal correctly but misses parts of the path in between. Specifically, the corridor leading up to the elevator, visible in the overt run (a) is missing in the internal simulation (b). While the first 15 s of the covert run appears similar to its overt counterpart, the total time is much shorter, 18 s compared to 29 s. Hence, the shorter simulation time is due to a lost segment of the simulation rather than a general increase in execution speed over the whole simulation. At t = 15 s, the robot is approaching a door leading to the corridor, followed by a left turn toward the elevator. An enlargement of this sequence of events, from t = 14 s to t = 20 s, is presented in **Figure 11C**. The door shows up in the figure as a narrow passage just before the left turn. It is likely that sensory interactions, when exiting the door and facing the corridor wall, are similar to perceptions when approaching the elevator. This hypothesis is confirmed by comparing sensor events from the covert condition to a subset of events from the overt case, when approaching the elevator (24 < t < 26 s). A period between 16.25 and 17 s from the covert case shows very similar sensor interactions to the selected period from the covert condition. This pinpoints the time where events from the door passage are confused with events form the elevator, and a segment for the original path disappears. **Figure 11D** presents a magnification of the covert condition, with laser scans prior to 16.5 s are colored in green and scans after 16.5 s are red. Green scans belong to the door passage, while red scans represent the elevator.

8. DISCUSSION

We present a robot model that can execute both overt and covert actions based on human demonstrations. The presented system, implemented on a simulated Kompaï robot (Robosoft, 2011), can learn from several demonstrations and execute a novel path through an apartment environment toward a goal. We also demonstrate that the system is able to generate internal simulations of sensory–motor experiences from executing a specific goal-directed behavior.

The model presents an associationist's approach to control an internal simulation, representing knowledge as coordination between perceptions and actions. Hence, despite the fact that the model is evaluated as a method for path following and generation, the system state is only represented implicitly, and very little application specific information is introduced. Models using *simulated experience* [c.f. Sutton and Barto (1998)] to improve valuations of explicitly represented system states exist in the form of reinforcement learning Dyna-based algorithms [e.g., Santos et al. (2012) and Lowe and Ziemke (2013)]. However, these algorithms are limited to updating (either randomly or heuristically) already experienced states and do not simulate novel paths. We use a morphologically simple robot, allowing us to study the principles of simulation in human-like environments without introducing the complexity of a humanoid robot. The selected platform [Microsoft (2015)] is freely available, facilitating replication of, and comparisons with, the present study.

The results provide support for Hypothesis 1 (Section 6). The system can generate the sensory–motor experiences of executing a novel path through the environment, without actually executing these actions. The robot model is able to pursue goals during both overt and covert behaviors. While the proportion of runs leading to a goal was slightly lower in the covert condition (82%), compared to 94% during overt action, the robot's ability to pursue the correct goal is significantly better during covert action.

One possible interpretation of this result is that internal simulation could potentially be beneficial as a training exercise since difficult skills are "easier" to execute covertly, increasing the likelihood of successful reenaction. Motor imagery and other forms of imagery have been used to increase the performance of athletes and for rehabilitation (Guillot and Collet, 2005; Munzert et al., 2009). Of particular interest to our experiment is the work of Vieilledent et al. (2003), who investigated the influence of mental imagery on path navigation. In their study, subjects were to navigate blindfolded three different 12.5-m long hexagonal shapes, indicated by wooden beams laid out on the floor. They found that a learning period, including either mental imagery, mental imagery, and simultaneous walking or walking with sensory feedback from a wooden beam, resulted in increased performance compared to a resting condition or walking without mental imagery.

In a similar study, Commins et al. (2013) did not find any increase in performance of mental imagery, but they did find that errors increased with distance in both the actual walking condition and imagery condition. Thus, more studies are needed to investigate the actual benefit of mental imagery in navigation. While it is possible that there are several factors that contribute to the differences in performance, our finding that goal pursuit is easier to execute covertly might be a clue to why mental training is advantageous in some cases.

With regard to the differences observed between the overt and covert runs, it should be noted that humans do not necessarily perform perfectly when acting based on internal simulations. Vieilledent et al. (2003) and Commins et al. (2013) showed that blind navigation resulted in similar trajectories and relatively accurate behaviors in terms of both deviation from target and temporal extension, but Vieilledent et al. (2003) found that in the blindfolded condition, the path was not as straight and turns where not as sharp leading to a more circular shape and also some distortions of the overall shape. From a simulation theory perspective, it would be suggested that even the blindfolded walking is based on chained simulations of covert actions and perceptions, but in this case guided by the additional proprioceptive feedback.

In light of these results, we should not expect the robot model to reproduce a perfect trajectory toward the target during covert action. This appears to be the case. We hypothesized (H2, Section 6) that successful internal simulations should have the same temporal extension as their overt counterpart. This hypothesis was not confirmed. The model generated internal simulations that were both longer and shorter than the overt counterparts, producing significantly different temporal distributions compared to overt results. Two cases were analyzed in detail: (1) indicating prolongation due to sensory–motor event oscillation
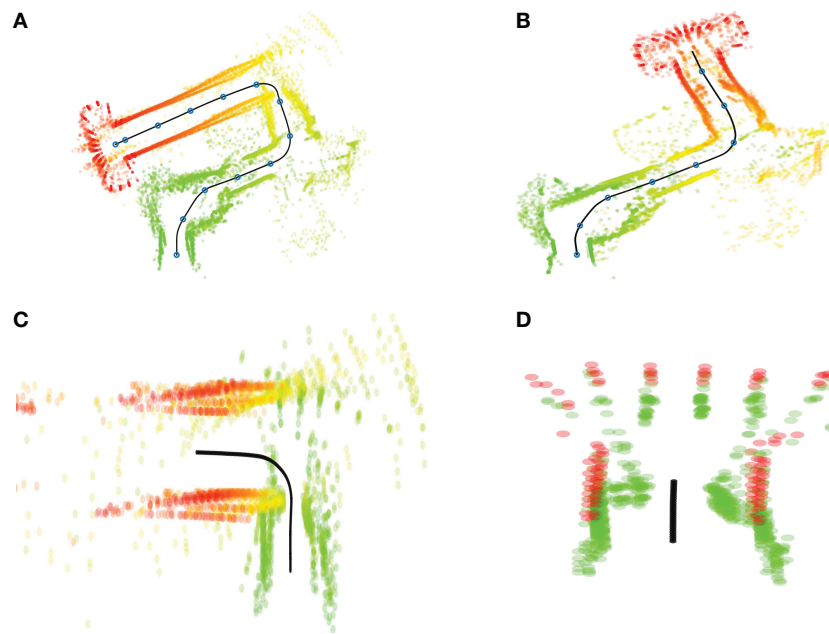
FIGURE 11 | Typical runs from condition Bed–Elevator. (A) Laser scans in relation to the path, executed overtly from bed (green) to the elevator (red). Blue circles along the path, represented by the black line, mark 2.5-s intervals. **(B)** Corresponding laser scans generated in covert mode. **(C)** Magnification of selected period (14–20 s) from the overt condition. **(D)** Corresponding (15–17 s) magnification from covert condition. See text for details.

caused by high model uncertainty (**Figure 10**) and (2) abbreviation caused by strong event similarities along the simulated path (**Figure 11**).

These results indicate that multiple types of distortions could affect internal simulations. If similar effects are present also during human mental imagery, we should be able to find longer simulation times in situations that are difficult for participants to reenact covertly. It is also possible that participants demonstrate shorter execution times during mental imagery in cases where it is easy for participants to mistake one location for another, causing parts of the path to be left out from the internal simulation. Both these,

and other, effects may appear simultaneously, and it may therefore be difficult to analyze mental imagery solely based on its total execution time.

AUTHOR CONTRIBUTIONS

The main content of the paper was written by EB and HS, with input from TZ and RL. Software implementations and experiments were executed by EB, including creations of figures. TZ and RL contributed with valuable background knowledge, edits, and shorter sections of text.

REFERENCES

Anderson, J. R. (1996). ACT: a simple theory of complex cognition. *Am. Psychol.* 51, 355. doi:10.1037/0003-066X.51.4.355

Anquetil, T., and Jeannerod, M. (2007). Simulated actions in the first and in the third person perspectives share common representations. *Brain Res.* 1130, 125–129. doi:10.1016/j.brainres.2006.10.091

Baldassarre, G. (2003). "Forward and bidirectional planning based on reinforcement learning and neural networks in a simulated robot," in *Anticipatory Behavior in Adaptive Learning Systems: Foundations, Theories, and Systems*, Vol. 2684, eds M. V. Butz, O. Sigaud, and P. Grard (Berlin; Heidelberg: Springer Verlag), 179–200.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660. doi:10.1017/S0140525X99532147

Barsalou, L. W. (2005). "Situated conceptualization," in *Handbook of Categorization in Cognitive Science*, eds H. Cohen and C. Lefebvre (Amsterdam: Elsevier Science Ltd.), 619–650.

Barsalou, L. W., Simmons, K. W., Barbey, A. K., and Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends Cogn. Sci.* 7, 84–91. doi:10.1016/S1364-6613(02)00029-3

Billing, E. A. (2012). *Cognition Rehearsed – Recognition and Reproduction of Demonstrated Behavior*. Ph.D. thesis, Umeå University, Department of Computing Science, Umeå.

Billing, E. A., and Hellström, T. (2010). A formalism for learning from demonstration. *Paladyn J. Behav. Robot.* 1, 1–13. doi:10.2478/s13230-010-0001-5

Billing, E. A., Hellström, T., and Janlert, L. E. (2010). "Behavior recognition for learning from demonstration," in *Proceedings of IEEE International Conference on Robotics and Automation* (Anchorage, AK: IEEE), 866–872.

Billing, E. A., Hellström, T., and Janlert, L. E. (2011a). "Predictive learning from demonstration," in *Agents and Artificial Intelligence: Second International Conference, ICAART 2010, Valencia, Spain, January 22–24, 2010. Revised Selected Papers*, eds J. Filipe, A. Fred, and B. Sharp (Berlin: Springer Verlag), 186–200.

Billing, E. A., Hellström, T., and Janlert, L. E. (2011b). "Robot learning from demonstration using predictive sequence learning," in *Robotic Systems – Applications, Control and Programming*, ed. A. Dutta (Rijeka: InTech), 235–250.

Billing, E. A., Hellström, T., and Janlert, L. E. (2015). Simultaneous recognition and reproduction of demonstrated behavior. *Biol. Inspired Cogn. Archit.* 12, 43–53. doi:10.1016/j.bica.2015.03.002

Brooks, R. A. (1991). Intelligence without representation. *Artif. Intell.* 47, 139–159. doi:10.1016/0004-3702(91)90053-M

Chrisley, R. (1990). "Cognitive map construction and use: a parallel distributed processing approach," in *Connectionist Models: Proceedings of the 1990 Summer School* (San Mateo, CA: Morgan Kaufman), 287–302.

Clancey, W. J. (1997). *Situated Cognition: On Human Knowledge and Computer Representations*. Cambridge: Cambridge University Press.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.

Colder, B. (2011). Emulation as an integrating principle for cognition. *Front. Hum. Neurosci.* 5:54. doi:10.3389/fnhum.2011.00054

Commins, S., McCormack, K., Callinan, E., Fitzgerald, H., Molloy, E., and Young, K. (2013). Manipulation of visual information does not change the accuracy of distance estimation during a blindfolded walking task. *Hum. Mov. Sci.* 32, 794–807. doi:10.1016/j.humov.2013.04.003

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: GP Putnam's Sons.

Demiris, J., and Hayes, G. R. (2002). "Imitation as a dual-route process featuring predictive and learning components: a biologically plausible computational model," in *Imitation in Animals and Artifacts*, eds K. Dautenhahn and C. L. Nehaniv (Cambridge, MA: MIT Press), 327–361.

Demiris, Y., Aziz-Zadeh, L., and Bonaiuto, J. (2014). Information processing in the mirror neuron system in primates and machines. *Neuroinformatics* 12, 63–91. doi:10.1007/s12021-013-9200-7

Demiris, Y., and Khadhouri, B. (2006). Hierarchical attentive multiple models for execution and recognition of actions. *Rob. Auton. Syst.* 54, 361–369. doi:10.1016/j.robot.2006.02.003

Demiris, Y., and Simmons, G. (2006). Perceiving the unusual: temporal properties of hierarchical motor representations for action perception. *Neural Netw.* 19, 272–284. doi:10.1016/j.neunet.2006.02.005

Fodor, J. A., and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition* 28, 3–71. doi:10.1016/0010-0277(88)90031-5

Fullér, R. (2000). *Introduction to Neuro-Fuzzy Systems*. Heidelberg: Physica-Verlag.

Gallese, V. (2003). The manifold nature of interpersonal relations: the quest for a common mechanism. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 517–528. doi:10.1098/rstb.2002.1234

Gallese, V. (2005). Embodied simulation: from neurons to phenomenal experience. *Phenomenol. Cogn. Sci.* 4, 23–48. doi:10.1007/s11097-005-4737-z

Gigliotta, O., Pezzulo, G., and Nolfi, S. (2010). "Emergence of an internal model in evolving robots subjected to sensory deprivation," in *From Animals to Animats 11*, eds S. Doncieux, B. Girard, A. Guillot, J. Hallam, J.-A. Meyer, and J.-B. Mouret (Berlin; Heidelberg: Springer Verlag), 575–586.

Gross, H. M., Heinze, A., Seiler, T., and Stephan, V. (1999). Generative character of perception: a neural architecture for sensorimotor anticipation. *Neural Netw.* 12, 1101–1129. doi:10.1016/S0893-6080(99)00047-7

Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behav. Brain Sci.* 27, 377–396. doi:10.1017/S0140525X04000093

Guillot, A., and Collet, C. (2005). Duration of mentally simulated movement: a review. *J. Mot. Behav.* 37, 10–20. doi:10.3200/JMBR.37.1.10-20

Harnad, S. (1990). The symbol grounding problem. *Physica D* 42, 335–346. doi:10.1016/0167-2789(90)90087-6

Haruno, M., Wolpert, D. M., and Kawato, M. (2003). "Hierarchical MOSAIC for movement generation," in *International Congress Series 1250* (Paris: Elsevier Science B.V.), 575–590.

Hesslow, G. (2002). Conscious thought as simulation of behaviour and perception. *Trends Cogn. Sci.* 6, 242–247. doi:10.1016/S1364-6613(02)01913-7

Hesslow, G. (2012). The current status of the simulation theory of cognition. *Brain Res.* 1428, 71–79. doi:10.1016/j.brainres.2011.06.026

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *J. Cogn. Neurosci.* 21, 1229–1243. doi:10.1162/jocn.2009.21189

Hiraki, K., Sashima, A., and Phillips, S. (1998). From egocentric to allocentric spatial behavior: a computational model of spatial development. *Adapt. Behav.* 6, 371–391. doi:10.1177/105971239800600302

Hoffmann, H., and Möller, R. (2004). "Action selection and mental transformation based on a chain of forward models," in *From Animals to Animats 8*, eds S. Schaal, A. Ijspeert, A. Billard, S. Vijayakumar, J. Hallam, and J.-A. Meyer (Cambridge, MA: MIT Press), 213–222.

Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage* 14, S103–S109. doi:10.1006/nimg.2001.0832

Jirenhed, D. A., Hesslow, G., and Ziemke, T. (2001). "Exploring internal simulation of perception in mobile robots," in *Proceedings of the Fourth European Workshop on Advanced Mobile Robotics*, eds K. O. Arras, A. J. Baerveldt, C.

Balkenius, W. Burgard, and R. Siegwart (Lund: Lund University Cognitive Studies), 107–113.

Jung, M., Hwang, J., and Tani, J. (2015). Self-organization of spatio-temporal hierarchy via learning of dynamic visual image patterns on action sequences. *PLoS ONE* 10:e0131214. doi:10.1371/journal.pone.0131214

Klir, G. J., and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. New Jersey: Prentice Hall.

K-Team. (2007). *Khepera Robot*. Available at: www.k-team.com

Lakoff, G., and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.

Lowe, R., and Ziemke, T. (2013). "Exploring the relationship of reward and punishment in reinforcement learning," in *IEEE Symposium on Adaptive Dynamic Programming And Reinforcement Learning (ADPRL)* (Singapore: IEEE), 140–147.

Mel, B. W. (1991). A connectionist model may shed light on neural mechanisms for visually guided reaching. *J. Cogn. Neurosci.* 3, 273–292. doi:10.1162/jocn.1991.3.3.273

Microsoft. (2015). *Microsoft Robotic Developer Studio*. Available at: http://www.microsoft.com/robotics/

Möller, R. (1999). "Perception through anticipation. A behaviour-based approach to visual perception," in *Understanding Representation in the Cognitive Sciences*, eds A. Riegler, M. Peschl, and A. von Stein (New York: Springer US), 169–176.

Munzert, J., Lorey, B., and Zentgraf, K. (2009). Cognitive motor processes: the role of motor imagery in the study of motor representations. *Brain Res. Rev.* 60, 306–326. doi:10.1016/j.brainresrev.2008.12.024

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., and Spivey, M. (2013). Computational grounded cognition: a new alliance between grounded cognition and computational modeling. *Front. Psychol.* 3:612. doi:10.3389/fpsyg.2012.00612

Pylyshyn, Z. W. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, MA: The MIT Press.

Robosoft. (2011). *Kompai Robot*. Available at: www.robosoft.com

Rumelhart, D. E., Smolensky, P., McClelland, J. L., and Hinton, G. E. (1986). "Schemata and sequential thought processes in PDP models," in *Parallel Distributed Processing*, eds J. L. McClelland, D. E. Rumelhart, and C. P. R. Group (Cambridge, MA: MIT Press), 7–57.

Santos, M., Martín, H. J. A., López, V., and Botella, G. (2012). Dyna-H: a heuristic planning reinforcement learning algorithm applied to role-playing game strategy decision systems. *Knowl. Based Syst.* 32, 28–36. doi:10.1016/j.knosys.2011.09.008

Siegwart, R., and Nourbakhsh, I. R. (2004). *Introduction to Autonomous Mobile Robots*. London: MIT Press.

Smith, S. M., Brown, H. O., Toman, J. E. P., and Goodman, L. S. (1947). The lack of cerebral effects of d-tubocurarine. *Anesthesiology* 8, 1–14. doi:10.1097/00000542-194701000-00001

Stening, J., Jacobsson, H., and Ziemke, T. (2005). "Imagination and abstraction of sensorimotor flow: towards a robot model," in *Proceedings of the Symposium on Next Generation Approaches to Machine Consciousness: Imagination, Development, Intersubjectivity and Embodiment (AISB'05)* (Hatfield: The University of Hertfordshire), 50–58.

Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Svensson, H. (2007). *Embodied Simulation as Off-line Representation*. Licentiate thesis, Department of Computer and Information Science, Institute of Technology, Linköping University, Linköping.

Svensson, H. (2013). *Simulations*. Ph.D. thesis, Department of Computer and Information Science, Institute of Technology, Linköping University, Linköping.

Svensson, H., Morse, A., and Ziemke, T. (2009a). "Representation as internal simulation: a minimalistic robotic model," in *Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society* (Austin, TX: Cognitive Science Society, Inc.), 2890–2895.

Svensson, H., Morse, A. F., and Ziemke, T. (2009b). "Neural pathways of embodied simulation," in *Anticipatory Behavior in Adaptive Learning Systems: From Psychological Theories to Artificial Cognitive Systems*, Vol. 5499, eds G. Pezzulo, M. V. Butz, O. Sigaud, and G. Baldassarre (Berlin; Heidelberg: Springer), 95–114.

Svensson, H., Thill, S., and Ziemke, T. (2013). Dreaming of electric sheep? Exploring the functions of dream-like mechanisms in the development of mental imagery simulations. *Adapt. Behav.* 21, 222–238. doi:10.1177/1059712313491295

Tani, J. (1996). Model-based learning for mobile robot navigation from the dynamical systems perspective. *IEEE Trans. Syst. Man Cybern. B Cybern.* 26, 421–436. doi:10.1109/3477.499793

Tani, J., Nishimoto, R., and Paine, R. W. (2008). Achieving "organic compositionality" through self-organization: reviews on brain-inspired robotics experiments. *Neural Netw.* 21, 584–603. doi:10.1016/j.neunet.2008.03.008

Varela, F., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge: MIT Press.

Vernon, D. (2014). *Artificial Cognitive Systems: A Primer*. London: The MIT Press.

Vernon, D., Beetz, M., and Sandini, G. (2015). Prospection in cognition: the case for joint episodic-procedural memory in cognitive robotics. *Front. Robot. AI* 2:19. doi:10.3389/frobt.2015.00019

Vieilledent, S., Kosslyn, S. M., Berthoz, A., and Giraudo, M. D. (2003). Does mental simulation of following a path improve navigation performance without vision? *Cogn. Brain Res.* 16, 238–249. doi:10.1016/S0926-6410(02)00279-3

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychol. Rev.* 20, 158–177. doi:10.1037/h0074428

Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science* 269, 1880–1882. doi:10.1126/science.7569931

Wolpert, D. M., and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Netw.* 11, 1317–1329. doi:10.1016/S0893-6080(98)00066-5

Yamashita, Y., and Tani, J. (2008). Emergence of functional hierarchy in a multiple timescale neural network model: a humanoid robot experiment. *PLoS Comput. Biol.* 4:e1000220. doi:10.1371/journal.pcbi.1000220

Ziemke, T., Jirenhed, D.-A., and Hesslow, G. (2005). Internal simulation of perception: a minimal neuro-robotic model. *Neurocomputing* 68, 85–104. doi:10.1016/j.neucom.2004.12.005

Zwaan, R. A. (2003). "The immersed experiencer: toward an embodied theory of language comprehension," in *Psychology of Learning and Motivation*, Vol. 44 (San Diego, CA: Academic Press), 35–62.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Prospection in cognition: the case for joint episodic-procedural memory in cognitive robotics

David Vernon[1]*, Michael Beetz[2] and Giulio Sandini[3]

[1] Interaction Lab, School of Informatics, University of Skövde, Skövde, Sweden, [2] Institute for Artificial Intelligence, University of Bremen, Bremen, Germany, [3] Department of Robotics, Brain and Cognitive Sciences, Istituto Italiano di Tecnologia, Genova, Italy

Prospection lies at the core of cognition: it is the means by which an agent – a person or a cognitive robot – shifts its perspective from immediate sensory experience to anticipate future events, be they the actions of other agents or the outcome of its own actions. Prospection, accomplished by internal simulation, requires mechanisms for both perceptual imagery and motor imagery. While it is known that these two forms of imagery are tightly entwined in the mirror neuron system, we do not yet have an effective model of the mentalizing network which would provide a framework to integrate declarative episodic and procedural memory systems and to combine experiential knowledge with skillful know-how. Such a framework would be founded on joint perceptuo-motor representations. In this paper, we examine the case for this form of representation, contrasting sensory-motor theory with ideo-motor theory, and we discuss how such a framework could be realized by joint episodic-procedural memory. We argue that such a representation framework has several advantages for cognitive robots. Since episodic memory operates by recombining imperfectly recalled past experience, this allows it to simulate new or unexpected events. Furthermore, by virtue of its associative nature, joint episodic-procedural memory allows the internal simulation to be conditioned by current context, semantic memory, and the agent's value system. Context and semantics constrain the combinatorial explosion of potential perception-action associations and allow effective action selection in the pursuit of goals, while the value system provides the motives that underpin the agent's autonomy and cognitive development. This joint episodic-procedural memory framework is neutral regarding the final implementation of these episodic and procedural memories, which can be configured sub-symbolically as associative networks or symbolically as content-addressable image databases and databases of motor-control scripts.

Introduction

The goal of this article is to argue the case of the use of joint episodic memory to facilitate prospection and goal-directed action in cognitive robotics. The article begins with insights from the biological sciences regarding the prospective nature of action, leading to a discussion of the role of memory in prospection, and the realization of prospection through internal simulation. This sets the scene

for the introduction of ideo-motor theory, *vis-à-vis* sensory-motor theory, and an explanation of the importance of joint perceptuo-motor representations. This is then followed by two examples of how these principles have been applied in cognitive architectures and an argument in favor of explicit perceptuo-motor memory – joint episodic-procedural memory – over perceptuo-motor mappings. We finish with a description of a simple proof-of-principle example implementation of joint episodic-procedural memory for overt attention.

The Goal-Directed and Prospective Nature of Action

Evidence from many different fields of research, including psychology and neuroscience, suggests that the movements of biological organisms are organized as actions and not reactions (von Hofsten, 2004). While reactions are elicited by earlier events, actions are initiated by a motivated subject, they are defined by goals, and they are guided by prospective information (Vernon et al., 2010). For example, when performing manipulation tasks or observing someone else performing them, subjects fixate on the goals and sub-goals of the movements not on the body parts, e.g., the hands or the objects (Johansson et al., 2001; Flanagan and Johansson, 2003). This happens only if a goal-directed action is implied. When showing the same movements without the context of an agent, subjects fixate the moving object instead of the goal.

Evidence from neuroscience also shows that the brain represents movements in terms of actions even at the level of neural processes [see Vernon et al. (2010), Chapter 4]. For example, the primate brain has two areas devoted to controlling movements: the premotor cortex and the motor cortex. The premotor cortex is the area of the brain that is active during motor planning and it influences the motor cortex which then executes the motor program comprising an action. The premotor cortex receives strong visual inputs from a region in the brain known as the inferior parietal lobule. These inputs serve a series visuomotor transformations for reaching (Area F4) and grasping (Area F5). Single neuron studies have shown that most F5 neurons code for specific goal-directed actions, rather than their constituent movements. Furthermore, several F5 neurons, in addition to their motor properties, respond also to visual stimuli. These are referred to as visuomotor neurons. The significance of this is that the premotor cortex of primates encodes actions (including implicit goals and expected states) and not just movements. The goal, therefore, is the fundamental property of the action rather than the specific motoric details of how it is achieved.

In primates, two classes of visuomotor neurons can be distinguished within area F5: canonical neurons and mirror neurons (Rizzolatti and Fadiga, 1998). The activity of both canonical and mirror neurons correlates with two distinct circumstances. In the case of canonical neurons, the same canonical neuron fires when a monkey sees a particular object and also when the monkey actually grasps an object with the same characteristic features. On the other hand, mirror neurons (Gallese et al., 1996; Rizzolatti et al., 1996; Rizzolatti and Craighero, 2004) are activated both when an action is performed and when the same or similar action is observed being performed by another agent. These neurons are

specific to the goal of the action and not the mechanics of carrying it out. So, for example, a monkey observing another monkey, or a human, reaching for a nut will cause mirror neurons in the premotor cortex to fire; these are the same neurons that fire when the monkey actually reaches for a nut. However, if the monkey observes another monkey making exactly the same movements but there is no nut present – there is no apparent goal of the reaching action – then the mirror neurons do not fire. Similarly, different motions that comprise the same goal-directed action will cause the same mirror neurons to fire. It is the action that matters: mirror neurons are not active if there is no explicit or implied goal. Since goals focus on the future, not the present, this again demonstrates the importance of prospection in action.

Finally, there is another reason why actions are guided by prospective information as opposed to instantaneous feedback data. Often, events in the agent's world may precede the feedback signals about them because the delays in the control pathways of biological systems may be substantial. If you cannot rely on feedback, the only way to overcome the problem is to anticipate what is going to happen next and to use that information to control one's behavior.

Prospection, then, is central to cognition. The question is how this prospection is achieved. The answer is, somewhat surprisingly, memory.

Memory

Memory facilitates the persistence of knowledge and forms a reservoir of experience. Without it, it would be impossible for the system to learn, develop, adapt, recognize, plan, deliberate, and reason (Vernon, 2014). Memory functions to preserve what has been achieved through learning and development, ensuring that, when a cognitive system adapts to new circumstances, it does not lose its ability to act effectively in situations to which it had adapted previously. But memory has another role in addition to preserving past experience: to anticipate the future. It forms the basis for one of the central pillars of cognitive capacity, i.e., the ability to simulate internally the outcomes of possible actions and select the one that seems most appropriate for the current situation. Viewed in this light, memory can be seen as a mechanism that allows a cognitive agent *to prepare to act*, overcoming through anticipation the inherent "here-and-now" limitations of its perceptual capabilities.

We can distinguish memory in many ways (Squire, 2004; Wood et al., 2012). For example, it can be distinguished on the basis of the nature of what is remembered and the type of access we have to it. Specifically, memory can be categorized as either *declarative* or *procedural*, depending on whether it captures knowledge of things – facts – or actions – skills. Sometimes they are characterized as memory of knowledge and know-how: "knowing that" and "knowing how."[1] This distinction applies mainly to long-term memory but short-term memory too has a declarative aspect. Declarative memory is sometimes referred to as *propositional memory* because it refers to information about the agent's world

---

[1] The distinction between *knowing that* and *knowing how* was made in 1949 by Gilbert Ryle in his book *The Concept of Mind* (Ryle, 1949)

that can be expressed in the form of propositions. This is significant because propositions are either true or false. Thus, declarative memory typically deals with factual information. This is not the case with skill-oriented procedural memory. As a consequence, declarative memories, in the form of knowledge, can be communicated from one agent to another through language, for example, whereas procedural memories can only be demonstrated.

Two different types of declarative memory can be distinguished. These are *episodic memory* and *semantic memory*. Episodic memory (Tulving, 1972, 1984) plays a key role in cognition and in the anticipatory aspect of cognition in particular. It refers to specific instances in the agent's experience while semantic memory refers to general knowledge about the agent's world which may be independent of the agent's specific experiences. In this sense, episodic memory is autobiographical. By its very nature in encapsulating some specific event in the agent's experience, episodic memory has an explicit spatial and temporal context: what happened, where it happened, and when it happened. This temporal sequencing is the only element of structure in episodic memory. Episodic memory is a fundamentally *constructive* process (Seligman et al., 2013). Each time an event is assimilated into episodic memory, past episodes are reconstructed. However, they are reconstructed a little differently each time. This constructive characteristic is related to the role that episodic memory plays in the process of internal simulation that forms the basis of prospection, the key anticipatory function of cognition.

In contrast, semantic memory "is the memory necessary for the use of language. It is a mental thesaurus, organized knowledge a person possesses about words and other verbal symbols, their meaning and referents, about relations among them, and about rules, formulas, and algorithms for the manipulation of the symbols, concepts, and relations."[2]

Episodic memory and semantic memory differ in many ways. In general, semantic memory is associated with how we understand (or model) the world around us, using facts, ideas, and concepts. On the other hand, episodic memory is closely associated with experience: perceptions and sensory stimulus. While episodic memory has no structure other than its temporal sequencing, semantic memory is highly structured to reflect the relationships between constituent concepts, ideas, and facts. Also, the validity (or truth, since semantic memory is a subset of propositional declarative memory) of semantic memories is based on social agreement rather than personal belief, as it is with episodic memory.[3] Semantic memory can be derived from episodic memory through a process of generalization and consolidation. Episodic memory can be both short-term and long-term while semantic memory and procedural memory are long-term.

Memory and Prospection

Memory plays at least four roles in cognition: it allows us to remember past events, anticipate future ones, imagine the

viewpoint of other people, and navigate around our world. All four involve *self-projection*: the ability of an agent to shift perspective from itself in the here-and-now and to take an alternative perspective. It does this by *internal simulation*, i.e., the mental construction of an imagined alternative perspective (Schacter et al., 2008). Thus, there are four forms of internal simulation (Buckner and Carroll, 2007):

1. Episodic memory (remembering the past).
2. Navigation (orienting yourself topographically, i.e., in relation to your surroundings).
3. Theory of mind (taking someone else's perspective on matters).
4. Prospection (anticipating possible future events).

Each form of simulation has a different orientation (past, present, or future) and each refers to the perspective of either the first person, i.e., the agent itself, or another person.

Prospection – the mental simulation of future possibilities – plays a central role in organizing perception, cognition, affect, memory, motivation, and action (Seligman et al., 2013). Prospection is referred to in various ways, e.g., *episodic future thinking*, *memory of the future*, *pre-experiencing*, *proscopic chronesthesia*, *mental time travel*, and just plain *imagination* and it can involve conceptual content and affective – emotional – states (Buckner and Carroll, 2007).

Recent evidence suggests that all four kinds of internal simulation involve a single core brain network and this network overlaps what is known as the *default-mode network*, a set of interconnected regions in the brain that is active when the agent is not occupied with some attentional tasks (Østby et al., 2012).

It is significant that all four forms of simulation are constructive, i.e., they involve a form of imagination. There is a difference between knowing about the future and projecting ourselves into the future. The latter is experiential and the former is not. Thus, episodic memory (memory of experiences) and semantic memory (memory of facts) facilitate different types of prospection. Episodic memory allows you to re-experience your past and pre-experience your future. There is evidence that projecting yourself forward in time is important when you form a goal, creating a mental image of yourself acting out the event and then episodically pre-experiencing the unfolding of a plan to achieve that goal. This use of episodic memory in prospection is referred to as *episodic future thinking*, a term coined by Cristina Atance and Daniela O'Neill to refer to the ability to project oneself forward in time to pre-experience an event (Atance and O'Neill, 2001; Szpunar, 2010).

The constructive aspect of episodic memory, whereby old episodic memories are reconstructed slightly differently every time a new episodic memory is assimilated or remembered, is particularly important in the context of internal simulation of events that have not been previously experienced. While episodic memory certainly needs some constructive capacity to assemble individual details into a coherent memory of a given episode, the *constructive episodic simulation hypothesis* (Schacter and Addis, 2007a,b; Schacter et al., 2008; Szpunar, 2010) suggests that its role in prospection involving the simulation of multiple possible futures imposes an even greater need for a constructive capacity because of the need to extrapolate beyond past experiences. In

---

[2]This quotation explaining the characteristics of semantic memory appears in Endel Tulving's 1972 article (Tulving, 1972), p. 386 and is quoted in his Précis (Tulving, 1984). While this definition of semantic memory dates from 1972, it is still valid today. It also explains the linguistic origins of the term.

[3]Semantic memory and episodic memory can be contrasted in many other ways: twenty-seven differences are listed in Tulving (1983), p. 35.

other words, simulating multiple yet-to-be-experienced futures requires flexibility in episodic memory. This flexibility is possible because episodic memory is not an exact and perfect record of experience but one that conveys the essence of an event and is open to re-combination.

It is also significant that when humans imagine the future, they not only anticipate an event, but they also anticipate how they feel about that event. These are referred to as *hedonic* consequences of the event: whether we feel good about it or bad about it, whether it is associated with pleasure or pain, and lack of concern or fear. Thus, the pre-experience of prospection also involves "pre-feeling." The brain accomplishes prospection by simulating the event and the associated hedonic experience (Gilbert and Wilson, 2007). While pre-feeling is not always reliable because contextual factors also play a part in the hedonic experience, this hedonic aspect of episodic memory is important because it reflects the affective nature of cognition and opens up a plausible way to factor emotional drives and value systems into the operation of memory, prospection, and action selection.
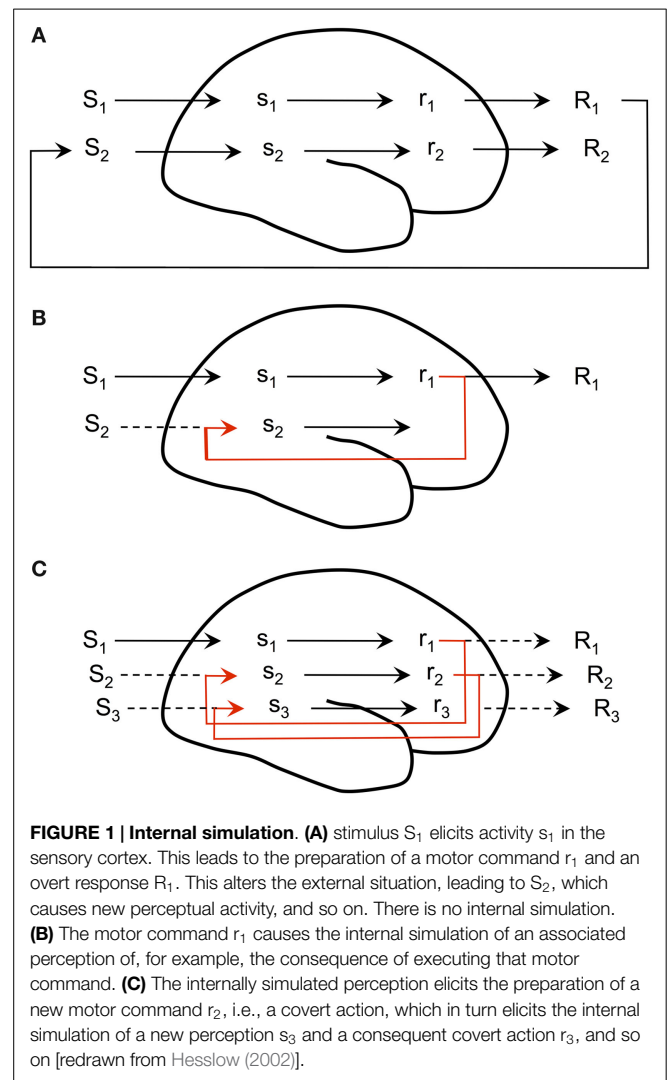
Internal Simulation and Action

In the preceding section, we considered internal simulation entirely in terms of memory-based self-projection, using re-assembled combinations of episodic memory to pre-experience possible futures, re-experience (and possibly adjust) past experiences, and project ourselves into the experiences of others. However, we know that action plays a significant role in our perceptions so the question then is: does action play a role in internal simulation? The answer is a clear "yes" (Hesslow, 2002, 2012; Svensson et al., 2007). Internal simulation extends beyond episodic memory and includes simulated interaction, particularly embodied interaction. Although the terms simulation, internal simulation, and mental simulation are widely used, you will also see references being made to *emulation*, very often when the approach endeavors to model the exact mechanism by which the simulation is produced (Grush, 2004).

The Simulation Hypothesis

There are a number of simulation theories, but perhaps the most influential is what is known as the *simulation hypothesis* (Hesslow, 2002, 2012). It makes three core assumptions:

1. The regions in the brain which are responsible for motor control can be activated *without* causing bodily movement.
2. Perceptions can be caused by internal brain activity and not just by external stimuli.
3. The brain has associative mechanisms that allow motor behavior or perceptual activity to evoke other perceptual activity.

The first assumption allows for simulation of actions and is often referred to as *covert action* or *covert behavior*. The second allows for simulation of perceptions. The third assumption allows simulated actions to elicit perceptions that are like those that would have arisen if the actions had actually been performed. There is an increasing amount of neurophysiological evidence in support of all three assumptions (Svensson et al., 2013). If we link these assumptions together, we see that the simulation hypothesis

FIGURE 1 | Internal simulation. **(A)** stimulus S_1 elicits activity s_1 in the sensory cortex. This leads to the preparation of a motor command r_1 and an overt response R_1. This alters the external situation, leading to S_2, which causes new perceptual activity, and so on. There is no internal simulation. **(B)** The motor command r_1 causes the internal simulation of an associated perception of, for example, the consequence of executing that motor command. **(C)** The internally simulated perception elicits the preparation of a new motor command r_2, i.e., a covert action, which in turn elicits the internal simulation of a new perception s_3 and a consequent covert action r_3, and so on [redrawn from Hesslow (2002)].

shows how the brain can simulate extended perception-action-perception sequences by having the simulated perceptions elicit simulated action which in turn elicits simulated perceptions, and so on. **Figure 1** summarizes the simulation hypothesis, showing three situations, one where there is no internal simulation, one where a motor response to an input stimulus causes the internal simulation of an associated perception, and one where this internally simulated perception then elicits a covert action which in turn elicits a simulated perception and a consequent covert action, and so on.

Motor, Visual, and Mental Imagery

Action-directed internal simulation involves three different types of anticipation: implicit, internal, and external (Svensson et al., 2009). *Implicit anticipation* concerns the prediction of motor commands from perceptions (which may have been simulated in a previous phase of internal simulation). *Internal anticipation* concerns the prediction of the proprioceptive consequences of carrying out an action, i.e., the effect of an action on the agent's own body. *External anticipation* concerns the prediction of the consequences for external objects and other agents of carrying

out an action.[4] Implicit anticipation selects some motor activity (possibly covert, i.e., simulated) to be carried out based on an association between stimulus and actions; internal anticipation and external anticipation then predict the consequences of that action. Collectively, they simulate actions and the effects of actions.

Covert action involves what is referred to as *motor imagery* and simulation of perception is often referred to as *visual imagery*. Perceptual imagery would perhaps be a better term since there is evidence that humans use imagery from all the senses. In a way, motor imagery is also a form of perceptual imagery, in the sense that it involves the proprioceptive and kinesthetic sensations associated with bodily movement. However, reflecting the interdependence of perception and action, covert action often has elements of both motor imagery and visual imagery and, *vice versa*, the simulation of perception often has elements of motor imagery. Visual imagery and motor imagery are sometimes referred to collectively as *mental imagery* (Wintermute, 2012). Moulton and Kosslyn (2009) identify several different types of perceptual imagery and distinguish between two different types of simulation: *instrumental simulation* and *emulative simulation*. The former concerns itself only with the content of the simulation while the latter also replicates the process by which that content is created in the simulated event itself. They refer to this as *second-order simulation*.

Joint Perceptuo-Motor Representations

In the foregoing, we remarked on the fact that mental imagery, viewed as another way of expressing the process of internal simulation, comprises both visual imagery (or perceptual imagery) and motor imagery. More importantly, though, we noted that these two forms of imagery are tightly entwined: they complement each other and the simulation of perception and covert action both involve elements of visual and motor imagery.

Classical treatments of memory usually maintain a clear distinction between declarative memory and procedural memory, in general, and between episodic memory and procedural memory, in particular. However, contemporary research takes a slightly different perspective, binding the two more closely, e.g., the mirror neuron system, in particular. While it is still a major challenge to understand how these two memory systems are combined, this coupling is the basic idea underpinning joint perceptuo-motor representations: representations that bring together the motoric and sensory aspects of experience in one framework, such as that anticipated in the simulation hypothesis.

In this section, we look at four approaches that have been developed to address joint perceptuo-motor representations. First, we look at two approaches to implementing ideo-motor theory in cognitive robotics: Shanahan's Global Workspace Theory architecture and Demiris's HAMMER architecture. We follow this by highlighting two additional approaches that endeavor to integrate perceptuo-motor representations more tightly: the Theory of Event Coding (TEC) and Object-Action Complexes. Since none of these explicitly incorporate episodic or procedural memory, we then suggest a way of drawing the principal ideas of each together

in a form of explicit joint episodic-procedural memory. We then argue that this joint episodic-procedural memory allows several of the challenges of cognitive robotics to be addressed.

Before discussing these, to provide the necessary context for prospective perceptuo-motor representations, we first address the difference between sensory-motor theory and ideo-motor theory.

Sensory-Motor Theory and Ideo-Motor Theory

Broadly speaking, there are the two distinct approaches for planning actions: *sensory-motor* action planning and *ideo-motor* action planning (Stock and Stock, 2004). Sensory-motor action planning treats actions as reactive responses to sensory stimuli and assumes that perception and action use distinct and separate representational frameworks. The sensory-motor view builds on the classic unidirectional data-driven information-processing approach to perception, proceeding stage by stage from stimulus to percept and then to response. It is unidirectional in that it does not allow the results of later processing to influence earlier processing. In particular, it does not allow the resultant (or intended) action to impact on the related sensory perception.

Ideo-motor action planning, on the other hand, treats action as the result of internally generated goals. It is the idea of achieving some action outcome, rather than some external stimulus, that is at the core of how cognitive agents behave. This reflects the view of action described above, with action being initiated by a motivated subject, defined by goals, and guided by prospection. The key point of the ideo-motor principle is that the selection and control of a particular goal-directed movement depends on the anticipation of the sensory consequence of accomplishing the intended action: the agent images (e.g., through internal simulation) the desired outcome and selects the appropriate actions in order to achieve it.

There is an important difference, though, between the concrete movements comprising an action and the higher-order goals of an action. Typically, actors do not voluntarily preselect the exact movements required to achieve a desired goal. Instead, they select prospectively guided intention-directed goal-focused action, with the specific movements being adaptively controlled as the action is executed. Thus, ideo-motor theory should be viewed both as an anticipatory idea-centered way of selecting actions and as a way of bridging the higher-order conceptual representations of intentions and goals[5] with the concrete adaptive control of movements when executing that action (Ondobaka and Bekkering, 2012).

In contrast to sensory-motor models, ideo-motor theory assumes that perception and action share a common representational framework. Because ideo-motor models focus on goals, and because they use a common joint representation that embraces both perception and action, they provide an intuitive explanation of why cognitive agents, humans in particular, are so adept at and predisposed to imitation (Iacoboni, 2009). The essential idea is that when I see somebody else's (goal-directed) actions and

---

[4]The terms *internal anticipation* and *external anticipation* are also referred to as *bodily anticipation* and *environmental anticipation* (Svensson et al., 2013).

[5]Michael Tomasello and colleagues note that the distinction between intentions and goals is not always clearly made. Taking their lead from Michael Bratman (1998), they define an intention as a plan of action an agent chooses and commits itself to in pursuit of a goal. An intention therefore includes both a means (i.e. an action plan) as well as a goal (Tomasello et al., 2005).

the consequences of these actions, the representations of my own actions that would produce the same consequences are activated.
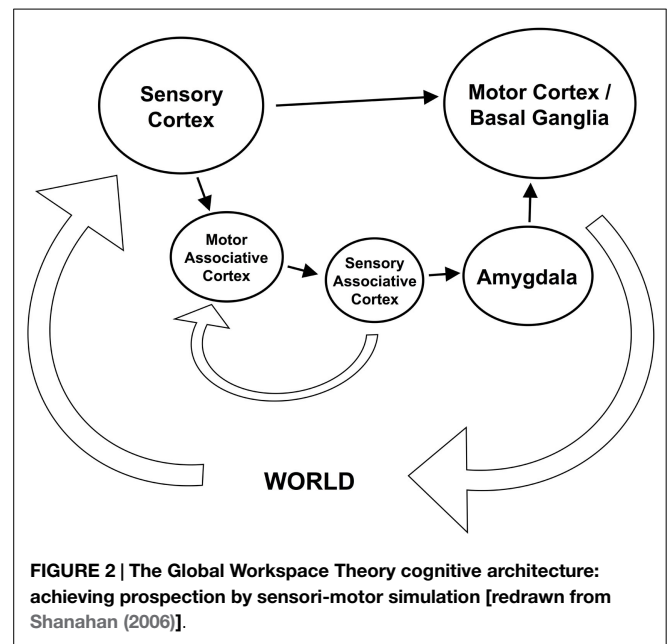
At first glance, ideo-motor theory seems to present a puzzle: how can the goal, achieved through action, cause the action in the first place? In other words, how can the later outcome affect the earlier action? This seems to be a case of *backward causation*. The solution to the puzzle is prospection. It is the anticipated goal state, not the achieved goal state, that impacts on the associated planned action. Goal-directed action, then, is a center-piece of ideo-motor theory, which is also referred to as the *goal trigger hypothesis* (Hommel et al., 2001).

Before proceeding to consider two cognitive architectures that build on ideo-motor theory, we mention *cognitive maps* to highlight the importance of joint perceptuo-motor representations in animal and robot cognition. The idea of a cognitive map was introduced by Tolman as a geometric representation to support navigation in biological agents (Tolman, 1948). While there is a certain lack of consensus on what exactly constitutes a cognitive map (Bennett, 1996; Eichenbaum et al., 1999), most agree that it involves metric information rather than purely topological information to encode spatial relationships in an allocentric framework and that it exploits path integration, at least partially, to effect navigation (Gallistel, 1989, 1990; Stachenfeld et al., 2014); for an alternative perspective, see Gaussier et al. (2002). In any case, a cognitive map combines memories of environmental cues (or perceptual landmarks) with geometrical properties of space that are specified by the remembered landmarks (Metta et al., 2010). Based on the existence of the hippocampus place cells (O'Keefe, 1976), O'Keefe and Nadel suggested that the hippocampal formation provides the neural basis for the cognitive map (O'Keefe and Nadel, 1978).

However, the hippocampus does not just create and store cognitive maps but it also plays a part in episodic memory, e.g., helping to minimize the similarities between new representations and representations that already exist in memory (McNaugton et al., 2006). As with episodic memory, it is also responsible for associating information in ways that allow flexible use of past experiences to guide future actions (*flexible memory expression)* (Eichenbaum et al., 1999; McNamara and Shelton, 2003). Furthermore, it has a role as a prediction mechanism for novelty detection and especially as a way to merge planning and sensory-motor function in a single coherent system (Gaussier et al., 2002). As McNaughton et al. note, ". our current understanding of [the hippocampal formation] underscores the growing paradigm shift in the neurosciences away from thinking about neural coding as being driven primarily by bottom-up, sensory inputs, but rather as a reflection of rich and complex internal dynamics" (McNaugton et al., 2006). Taken together, the characteristics of cognitive maps and the operation of the hippocampal formation echo the arguments being put forward in this paper about the importance of joint perceptuo-motor representations in cognition.

The Global Workspace Cognitive Architecture

Shanahan (Shanahan, 2005a,b, 2006; Shanahan and Baars, 2005) proposes a biologically plausible brain-inspired neural-level cognitive architecture in which cognitive functions such as anticipation and planning are realized through internal simulation of interaction with the environment. Action selection, both actual

FIGURE 2 | The Global Workspace Theory cognitive architecture: achieving prospection by sensori-motor simulation [redrawn from Shanahan (2006)].

and internally simulated, is mediated by affect. The architecture is based on an external sensori-motor loop and an internal sensori-motor loop in which information passes through multiple competing cortical areas and a global workspace (Baars, 1998, 2002).

Shanahan's cognitive architecture is comprised of the following components: a first-order sensori-motor loop, closed externally through the world, and a higher-order sensori-motor loop, closed internally through associative memories (see **Figure 2**). The first-order loop comprises the sensory cortex and the basal ganglia (controlling the motor cortex), together providing a reactive action-selection sub-system. The second-order loop comprises two associative cortex elements which carry out off-line simulations of the system's sensory and motor behavior, respectively. The first associative cortex simulates a motor output while the second simulates the sensory stimulus expected to follow from a given motor output. The higher-order loop effectively modulates basal ganglia action selection in the first-order loop via an affect-driven amygdala component. Thus, this cognitive architecture is able to anticipate and plan for potential behavior through the exercise of its "imagination" (*i.e.,* its associative internal sensori-motor simulation).

The HAMMER Architecture

While internal simulation is an essential aspect of human cognition, it is also an increasingly important part of artificial cognitive systems. For example, The Hierarchical Attentive Multiple Models for Execution and Recognition (HAMMER) architecture (Demiris and Khadhouri, 2006; Demiris et al., 2014) builds on the simulation hypothesis, accomplishing internal simulation using forward and inverse models which encode internal sensori-motor models that the agent would utilize if it were to execute that action (see **Figure 3**).

HAMMER deploys several inverse-forward pairs to simulate multiple possible futures using a winner-take-all attention process to select the most appropriate action to execute. HAMMER
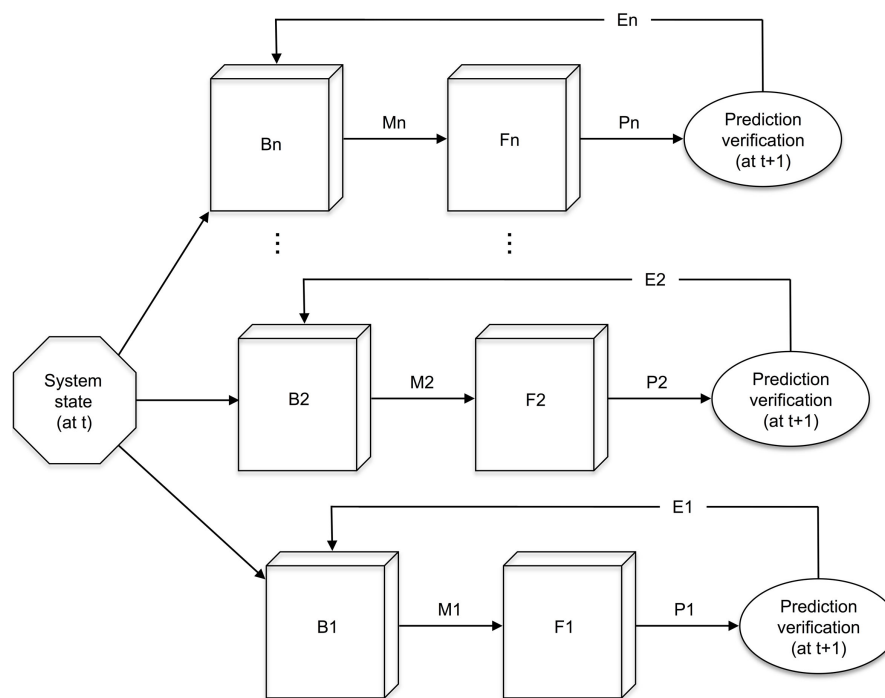
FIGURE 3 | The HAMMER architecture, showing multiple inverse models (B1 to Bn) taking as input the current system state, which includes a desired goal, suggesting motor commands (M1 to Mn), with which the corresponding forward models (F1 to Fn) form predictions of the system's next state (P1 to Pn). These predictions are verified at the next time state, resulting in a set of error signals (E1 to En). Redrawn from Demiris and Khadhouri (2006). See also Demiris et al. (2014) for an alternative rendering of the HAMMER architecture.

includes recurrent connections, thereby allowing multi-stage extended internal simulation and mental rehearsal. This provides the architecture with a way of encapsulating the internal simulation hypothesis proposed by Hesslow (2002, 2012).

The inverse model takes as input information about the current state of the system and the desired goal, and it outputs the motor commands necessary to achieve that goal. The forward model acts as a predictor. It takes as input the motor commands and simulates the perception that would arise if this motor command were to be executed, just as the simulation hypothesis envisages. HAMMER then provides the output of the inverse model as the input to the forward model. This allows a goal state (demonstrated, for example, by another agent or possibly recalled from episodic memory) to elicit the simulated action required to achieve it. This simulated action is then used with the forward model to generate a simulated outcome, i.e., the outcome that would arise if the motor commands were to be executed. The simulated perceived outcome is then compared to the desired goal perception and the results are then fed back to the inverse model to allow it to adjust any parameters of the action.

A distinguishing feature of the HAMMER architecture is that it operates multiple pairs of inverse and forward models in parallel, each one representing a simulation – a hypothesis – of how the goal action can be achieved. The choice of inverse/forward model pair is made by an internal attention process based on how close the predicted outcome is to the desired one. Furthermore, it provides for the hierarchical composition of primitive actions into more complex sequences.
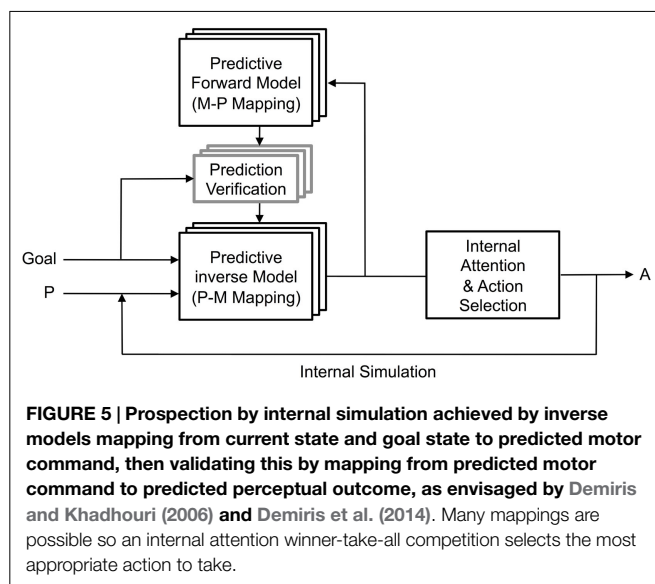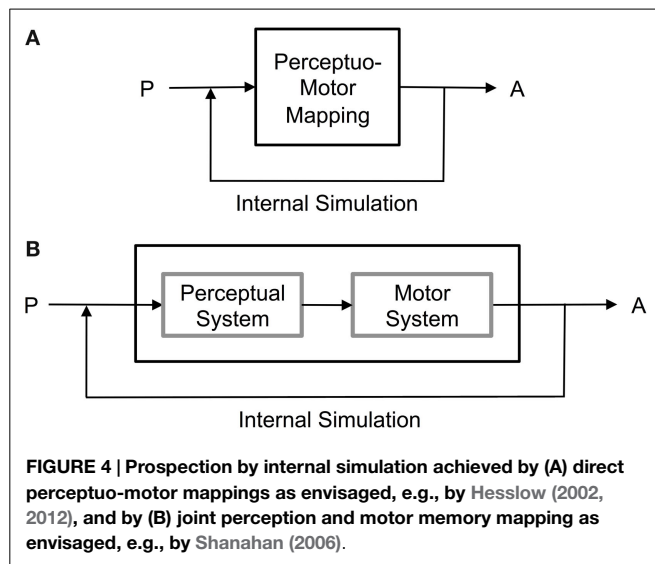
From Perceptuo-Motor Mappings to Perceptuo-Motor Memory

Both Global Workspace Theory and HAMMER are good models of the simulation hypothesis for internal simulation as a vehicle for prospection in cognition. However, they focus on the mapping between perception and motor command, with memory being left implicit (see **Figures 4** and **5**).

Other models, such as the *Theory of Event Coding* (TEC) (Hommel et al., 2001) and *Object Action Complexes* (OACs) (Krüger et al., 2011) attempt to provide a tighter coupling of the perceptual and motor aspect in a joint perceptuo-motor representation.

The Theory of Event Coding (TEC) is a representational framework for combining perception and action planning. It focuses mainly on the later stages of perception and the earlier phases of action. As such, it concerns itself with perceptual features but not with how those features are extracted or computed. Similarly, it concerns itself with preparing actions – action planning – but not with the final execution of those actions and the adaptive control of various parts of the agent's body. The main idea is that perception, attention, intention, and action all work with a common representation and, furthermore, that action depends on both external and internal causes.

TEC provides a basis for combining both sensory-motor and ideo-motor action planning (Stock and Stock, 2004) and to be a joint representation that serves both sensory-stimulated action and prospective goal-directed action. The core concept in TEC is the *event code*. This is effectively a structured aggregation of distal features of an event in the agent's world. These *feature codes* can

FIGURE 4 | Prospection by internal simulation achieved by (A) direct perceptuo-motor mappings as envisaged, e.g., by Hesslow (2002, 2012), **and by (B) joint perception and motor memory mapping as envisaged, e.g., by** Shanahan (2006).

FIGURE 5 | Prospection by internal simulation achieved by inverse models mapping from current state and goal state to predicted motor command, then validating this by mapping from predicted motor command to predicted perceptual outcome, as envisaged by Demiris and Khadhouri (2006) **and** Demiris et al. (2014). Many mappings are possible so an internal attention winner-take-all competition selects the most appropriate action to take.

be relatively simple (e.g., color, shape, moving to the left, falling) or more complex, such as an affordance. Also, TEC feature codes can emerge through the agent's experience; they do not have to be pre-specified. A given TEC feature code is associated with both the sensory system and the motor system. Typically, a feature code is derived from several proximal sensory sources (sensory codes) and it contributes to several proximal motor actuators (motor codes). Each *event code* comprises several feature codes representing some event, be it a perceived event or a planned event. Feature codes associated with an event are activated both when the event is perceived and when it is planned. Because features can be elements of many event codes, the activation of a given feature effectively primes, i.e., predisposes, all the other events of which this feature is a component.

Inspired by the Theory of Event coding, an Object-Action Complex (OAC) (Krüger et al., 2011) is a triple, i.e., a unit with three components: (*E, T, M*). *E* is an "execution specification"

(effectively an action). *T* is a function that predicts how the attributes that characterize the current state of the agent's world will change if the execution specification is executed. Effectively, of *T* as a prediction of how the agent's perceptions will change as a result of carrying out the actions given by *E*. *M* is a statistical measure of the success of the OAC's past predictions. In this way, an OAC combines the essential elements of a joint representation – perception and action – with a predictor that links current perceived states and future predicted perceived states that would result from carrying out that action. To a large extent, an OAC models an agent's interaction with the world as it executes some motor program (this is referred to as a low-level control program *C P* in the OAC literature). For example, an OAC might encode how to grasp an object or push an object into a given position and orientation (usually referred to as the object pose). OACs can be learned and executed, and they can be combined into more complex representations of actions and their perceptual consequences.

To date, neither TEC nor OAC has been embedded in the more general internal simulation framework described above. So, it is proposed here that there is a strong case for making memory – episodic and procedural – more explicit and embedding them in an internal simulation framework (such as that envisaged in the simulation hypothesis, the GWT Architecture, and the HAMMER architecture) in a way that makes their links more explicit (such as that envisaged in TEC and OAC). We address such a possible framework on the next section.

A Network-Based Joint Episodic-Procedural Memory for Internal Simulation

The core idea being proposed is to unwind the temporal and causal relationships between specific perceptions and actions that are implicit in the mappings of, e.g., GWT and HAMMER, and make them explicit in a weighted network of associations between perceptions and actions, in the manner of TEC and OAC (see **Figure 6**). In doing so, it makes the input to the joint perceptuo-motor mapping explicit as perceptual episodic memories and motoric procedural memories (see **Figures 7** and **8**). In the case of episodic memory, this provides a way to include other modalities including affective or hedonic memories. Procedural memory operates associatively in their own right: such procedural memories are not static but are dynamic and adapt as the action is executed.

Furthermore, such a framework allows one to expose the mapping dynamics explicitly. This may have several advantages in, for example, cognitive development which focuses on extending the timescale of the agent's prospective capacity and expanding the agent's repertoire of actions. Specifically, development might be facilitated by adjusting and adapting the network structure – its topology and strength of connectivity – as a function of experiential learning, intrinsic value systems (Merrick, 2010), including those derived from autonomic self-maintenance (Bickhard, 2000), and affective homeostasis and allostasis (Sterling, 2004, 2012; Morse et al., 2008; Ziemke and Lowe, 2009).

The network model of joint episodic-procedural memory facilitates prospection in three senses: prospection by predicting the
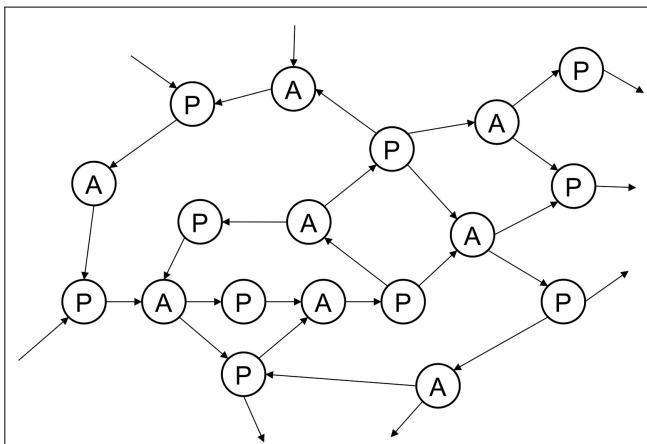
FIGURE 6 | Joint episodic-procedural memory as an explicit network of associations between perceptions and actions, drawn from episodic and procedural memories, unwinding the temporal and causal relationships between specific perceptions and actions that are implicit in the mappings of other perceptuo-motor representations.
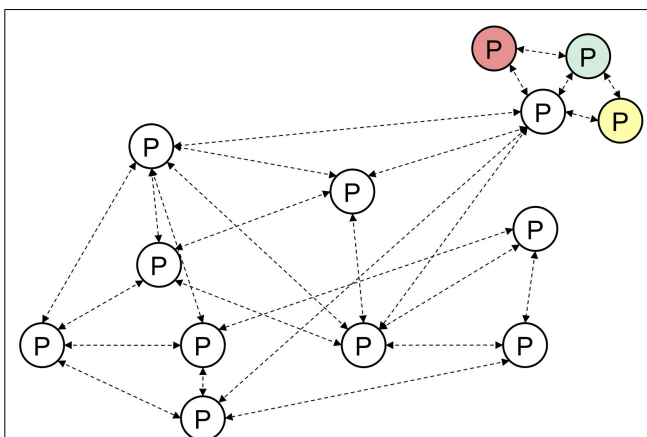

FIGURE 7 | The episodic elements of the joint episodic-procedural memory are drawn from episodic memory and therefore operate associatively in their own right. Furthermore, this provides a way to include other modalities of episodic memory (top right) including affective or hedonic memories.
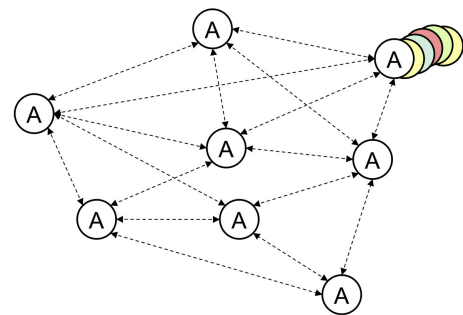

FIGURE 8 | The procedural elements of the joint episodic-procedural memory are drawn from procedural memory and, again, operate associatively in their own right. Such procedural memories are not static but are dynamic and adapt as the action is executed (top right).
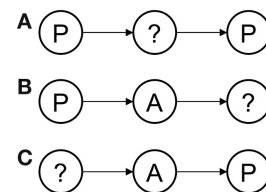

FIGURE 9 | The network model of joint episodic-procedural memory facilitates prospection in three senses: (A) prospection by predicting the outcome of an action carried out in given perceptual circumstances, (B) prospection by predicting the action required to achieve a goal in given perceptual circumstances, and (C) abductive inference of the perceptual states that explains an outcome of a give action.

outcome of an action carried out in given perceptual circumstances, prospection by predicting the action required to achieve a goal in given perceptual circumstances, and abductive inference of the perceptual states that explains an outcome of a give action (see **Figure 9**).

Keeping episodic memory explicit in this framework preserves the flexibility for adaptive reconstruction and novel association. Since episodic memory operates by recombining imperfectly recalled past experience, this allows it to simulate new or unexpected events as outlined above.

There is, however, a potential problem in that the scope for exponential growth in association is significant. Something is needed to constrain this potential combinatorial explosion if such a joint episodic-procedural memory system is to be capable of

useful prospection through internal simulation. Because the associative links are exposed explicitly in the network organization, this framework for a joint episodic-procedural memory allows the internal simulation to be conditioned by current context, semantic memory, and the agent's value system by adjusting the associative links. Context and semantics constrain the combinatorial explosion of potential perception-action associations and allow effective action selection in the pursuit of goals, while the value system modulates the memory network to promote the agent's autonomy and cognitive development.

Finally, the approach being suggested here is an abstract schema and is therefore neutral regarding the final implementation of these episodic and procedural memories. These can be effected either as an emergent cognitive system, instantiating them subsymbolically in a biologically inspired manner as associative networks [e.g., Hopfield nets such as in Mohan et al. (2014) or brain-based devices such as in Krichmar and Edelman (2005, 2006)]. Alternatively, they can be effected symbolically as more traditional AI systems. For example, episodic memory might be implemented using content-addressable image databases with traditional image indexing and recall algorithms, while procedural memory could be encapsulated in databases of motor-control scripts derived from experiential learning or from shared resources [e.g., Tenorth and Beetz (2009) and Tenorth et al. (2012, 2013)]. The traditional AI implementation, for the purpose of practical cognitive robotics, has a number of advantages.

Although episodic memory will typically exploit by iconic representations, these representations are often augmented by symbolic tags when derived from on-line repositories. This symbolic tagging makes the integration of semantic knowledge much easier. The fact that both episodic memory and procedural memory are derived from experience, directly or indirectly, also finesses the symbol grounding problem (Harnad, 1990; Sloman). The traditional AI implementation also renders the knowledge contained in the memory inherently transferrable to other agents, provided their sensory systems are compatible and there is a known mapping – direct or indirect – between the embodiments of each agent, as described in Argall et al. (2009).

An Example Joint Episodic-Procedural Memory for Overt Attention

The iCub is a 53 degree-of-freedom humanoid robot (see **Figure 10**) that was designed to be an open-systems platform for research in cognitive development (Sandini et al., 2007; Tsagarakis et al., 2007; Metta et al., 2010). It is approximately 1 m tall, weighs 22 kg, has visual, vestibular, auditory, and haptic sensors, and is capable of dexterous manipulation. To date, iCubs have been delivered to over 20 research laboratories in Europe and one in the U.S.A.[6]

The original iCub cognitive architecture (Sandini et al., 2007; Vernon et al., 2010) focused on gaze-modulated goal-directed reaching and locomotion. Episodic memory and procedural memory were designed to effect internal simulation in order to provide capabilities for prediction and model construction bootstrapped by learned affordances. Motivations encapsulated in the system's affective state addressed curiosity and experimentation, both of which are exploratory motives, triggered by exogenous and endogenous factors, respectively. This distinction between the exogenous and the endogenous was reflected in the overt attention system that could be triggered by both external and internal events. A simple process of homeostatic self-regulation governed by the affective state provided elementary action selection. Finally, all the various components of the cognitive architecture operated concurrently so that a sequence of states representing cognitive behavior emerges from the interaction of many separate parallel processes rather than being dictated by some pre-programed state-machine.

In the variant of the iCub cognitive architecture presented here, the separate episodic and procedural memories have been replaced by a simple proof-of-principle joint episodic-procedural memory (see **Figure 11**). This is the focus of the current article and the specific objective is to investigate how a joint episodic-procedural memory can be used for representation, development, and adaptation of scan-path patterns that result from overt and covert attention. This particular model of attention uses an information-theoretic saliency map (Bruce and Tsotsos, 2009) with an overt attention system comprising (1) the winner-take-all process effected by a selective tuning model to identify a single focus of attention (Tsotsos et al., 1995; Tsotsos, 2006, 2011), (2) an Inhibition-Of-Return (IOR) mechanism to attenuate the attention
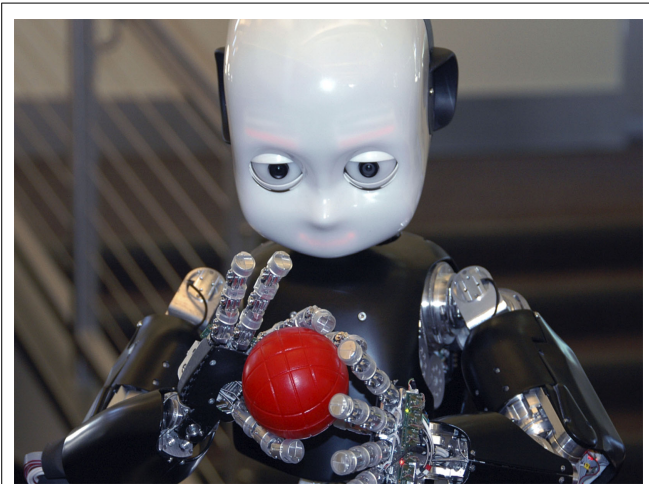

FIGURE 10 | The iCub humanoid robot: an open-systems platform for research in cognitive development.

value of previous winning locations so that new regions become the focus of attention, and (3) a habituation process to reduce the salience of the current focus of attention with time thereby ensuring that attention is fixated on a given point for a limited period (Zaharescu et al., 2004). Fixation points are represented using retinotopic images rather than conventional rectangular regularly sampled images. The retinotopic images are constructed using a scale and rotation-invariant log-polar transform (Braccini et al., 1981; Berton, 2006; Berton et al., 2006; Traver and Bernardino, 2010) to map the Cartesian camera image data to a non-uniformly sampled image that reflects the foveated sampling in the primate retina. The resultant scan path patterns are captured in an elementary joint episodic-procedural memory: the episodes are retinotopic log-polar images of the fixation points and the actions are the saccade angles.

The episodic memory in the iCub cognitive architecture is a simple associatively recalled memory of autobiographical events. It is a form on one-shot learning and does not generalize multiple instances of an observed event. In the current implementation, the episodic memory provides a purely visual iconic memory of landmark appearance using scale- and rotation-invariant[7] retinotopic log-polar images as the landmark representation (Braccini et al., 1981; Berton, 2006; Berton et al., 2006; Traver and Bernardino, 2010) with image recognition being effected using color histogram intersection (Swain and Ballard, 1990, 1991). In essence, the iCub episodic memory implements a form of content-addressable memory which is populated by log-polar landmark images acquired under the control of the iCub's covert and overt attention sub-system.

Procedural memory maintains a very simple repository of elementary actions. The current implementation comprises gaze motor commands in a body-centered frame of reference and symbolic tags denoting one of five possible associated actions (reach, push, grasp, locomote, or wait). These are just placeholders for

---

[6]For more information on the iCub robot see http://www.icub.org.

[7]The rotation invariance of log-polar images is restricted to roll: rotation about the camera's principal axis.
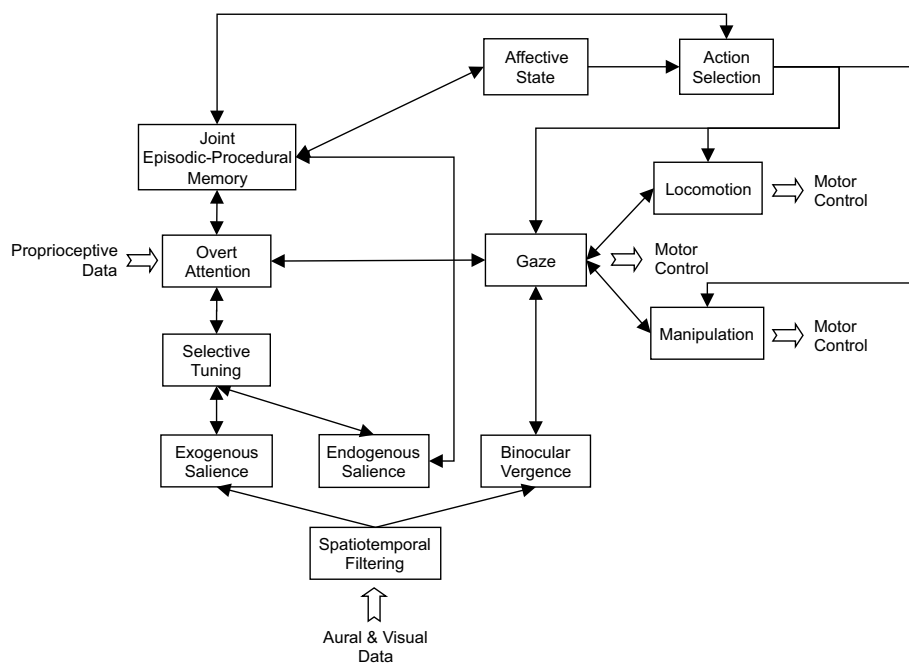
FIGURE 11 | A variant of the iCub cognitive architecture (Vernon et al., 2007, 2010**) targeting visual attention with information-theoretic exogenous salience (**Bruce and Tsotsos, 2009**), the Selective Tuning Model for** saccade selection (**Tsotsos et al., 1995; Tsotsos, 2006, 2011**), overt attention with inhibition of return and habituation modulated scan path dynamics (**Zaharescu et al., 2004**), and joint episodic-procedural memory.**

more flexible and adaptive gaze-directed motor control schemes [e.g., Lukic et al. (2012)] to be implemented later.

The joint episodic-procedural memory itself is a network of associations between motor events and pairs of sensory events. In this variant of the iCub cognitive architecture, a sensory event is a visual landmark which has been acquired by the iCub and stored in the episodic memory. A motor event is a gaze saccade with an optional reaching, grasping, or locomotion movement. Thus, joint episodic-procedural memory can be viewed as a directed network with two types of nodes, one representing sensory patterns – retinotopic log-polar images of the fixation points – and the other representing motor patterns – the saccade motor commands. A path through the network traverses alternately sensory and motor nodes and any clique in this memory network effectively captures a causal relationship between a sensory state, a motor state, and a subsequent sensory state (or a sequence of such associations). An extended path in this memory captures the scan path pattern of the robot as it pays attention to its visual environment (see **Figure 12**).

The key feature of this form of joint episodic-procedural memory representation of the attention pattern of the robot is that it lends itself to development: modulation or dynamically reconfiguration of the connectivity of this network – which is learned from experience – so that its prospective capacity increases as new memories are added as a result of the agent's interaction with its environment. Various forms of adaptive reconfiguration are currently being examined, some based on small world networks (Watts and Strogatz, 1998; Newman, 2000; Bohland and Minai, 2001; Kleinberg, 2006; Telesford et al., 2011) and others based

on information theoretic models that dynamically modulate the pathways in flow networks (Ulanowicz, 2000).

Conclusion

While action and prospection are intimately linked, most research on prospection has tended to focus on the constructive role of episodic memory (Tulving, 1972, 1984; Seligman et al., 2013), i.e., the so-called episodic future thinking (Atance and O'Neill, 2001), often achieved through internal simulation, i.e., the mental construction of an imagined alternative perspectives (Buckner and Carroll, 2007) and simulated embodied interaction (Svensson et al., 2007). Although hedonic affective experience has been addressed to some extent (Gilbert and Wilson, 2007; Lowe and Ziemke, 2011), procedural memory has been neglected in modeling prospective capacities. When it is included, it usually takes the form of distinct forward models that predict the sensory outcome of a given motor command (Hesslow, 2002, 2012; Shanahan, 2006) and inverse models that determine the action required to produce a given goal perception (Demiris and Khadhouri, 2006). Ideo-motor theory (Stock and Stock, 2004; Iacoboni, 2009) is an exception to this. It assumes that perception and action share a common representational framework and that action is the causal result of internally generated goals. Such a joint representation provides greater flexibility in prospection through both inductive inference and abductive inference.

With few exceptions, such as the Theory of Event Coding (Hommel et al., 2001) and object-action complexes
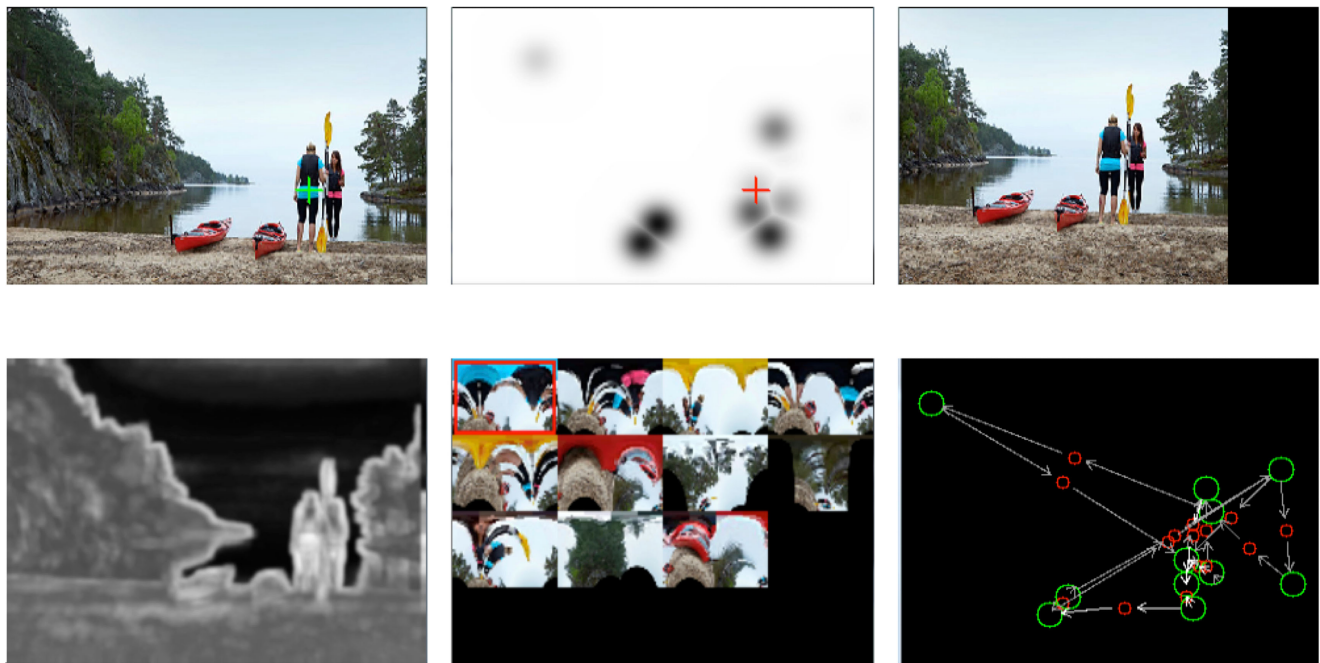
FIGURE 12 | A screen shot of an experiment using joint episodic-procedural memory with covert attention: (top left) the fixation point identified by the Selective Tuning Model (Tsotsos et al., 1995**; Tsotsos, 2006, 2011) based on (bottom left) the information-theoretic exogenous salience (**Bruce and Tsotsos, 2009**) and (top middle) the inhibition of return and habituation Gaussian modulation functions; (bottom middle) the retinotopic log-polar episodic memory – the current fixation image is denoted** by the red rectangle and the blue shirt is clearly visible in the fovea; (top right) the input image shifts to place the fixation point at the center; (bottom right) a graphic visualization of the joint episodic-procedural memory, with fixation-point episodes rendered as green circles, saccade actions rendered as red circles, and graph connections as directed arrows. Note that this graph is not registered with the image since the actions are specified in gaze angles, not image coordinates.

(Krüger et al., 2011), joint perceptuo-motor representations have received little attention and none have addressed integration of hedonic affective experience. Our conjecture is that an internal simulation capability founded on ideo-motor theory and joint representations, and drawing on recent progress in the modeling-related mirror neuron system (Gallese et al., 1996; Rizzolatti et al., 1996; Rizzolatti and Craighero, 2004; Thill et al., 2013), provides a viable way to approach the integration of procedural and episodic memory as a joint perceptuo-motor system. Our specific contention is that it is helpful to conceive of this joint episodic-procedural memory – for goal-directed internal simulation – as a network of associations between elements of both episodic and procedural memories. This perspective is neutral regarding the final implementation of these episodic and procedural memories and it can facilitate both emergent and cognitivist AI approaches.

We argue that such a framework meets several challenges in cognitive robotics such as the need to accommodate modal and modal episodic data and extended perceptuo-motor sequences, as well as mechanisms for conditioning the association dynamics with external constraints derived from semantic declarative knowledge, current context, and affective value signals. It also addresses the need to integrate the episodic and procedural knowledge gathered by robots as they operate of their physical environment with information extracted from web-based knowledge bases. This is particularly important if the power of indirect knowledge (acquired by interpreting third-party descriptions) is to be harnessed in the development of robot skills.

Acknowledgments

References

Argall, B. D., Chernova, S., Veloso, M., and Browning, B. (2009). A survey of robot learning from demonstration. *Rob. Auton. Syst.* 57, 469–483. doi:10.1016/j.robot. 2008.10.024

Atance, C. M., and O'Neill, D. K. (2001). Episodic future thinking. *Trends Cogn. Sci.* 5, 533–539. doi:10.1016/S1364-6613(00)01804-0

Baars, B. J. (1998). *A Cognitive Theory of Consciousness.* Cambridge: Cambridge University Press.

Baars, B. J. (2002). The conscious assess hypothesis: origins and recent evidence. *Trends Cogn. Sci.* 6, 47–52. doi:10.1016/S1364-6613(00)01819-2

Bennett, A. (1996). Do animals have cognitive maps. *J. Exp. Biol.* 199, 219–224.

Berton, F. (2006). "A brief introduction to log-polar mapping," in *Technical Report, LIRA-Lab,* (University of Genova).

Berton, F., Sandini, G., and Metta, G. (2006). Anthropomorphic visual sensors. *Encyclo. Sens.* 10, 1–16. doi:10.3389/fnsys.2014.00109

Bickhard, M. H. (2000). *Autonomy, function, and representation*. Available at: http://www.lehigh.edu/~mhb0/autfuncrep.html

Bohland, J. W., and Minai, A. A. (2001). Efficient associative memory using small-world architecture. *Neurocomputing* 3, 489–496. doi:10.1016/S0925-2312(01)00378-2

Braccini, C., Gambardella, G., and Sandini, G. (1981). A signal theory approach to the space and frequency variant filtering performed by the human visual system. *Signal Processing* 3, 231–240. doi:10.1016/0165-1684(81)90038-4

Bratman, M. E. (1998). "Intention and personal policies," in *Philosophical Perspectives*, ed. J. E. Tomberlin (Oxford: Blackwell), 3.

Bruce, N. D. B., and Tsotsos, J. K. (2009). Saliency, attention, and visual search: an information theoretic approach. *J. Vis.* 9, 1–24. doi:10.1167/9.3.5

Buckner, R. L., and Carroll, D. C. (2007). Self-projection and the brain. *Trends Cogn. Sci.* 11, 49–57. doi:10.1016/j.tics.2006.11.004

Demiris, Y., Aziz-Zahdeh, L., and Bonaiuto, J. (2014). Information processing in the mirror neuron system in primates and machines. *Neuroinformatics* 12, 63–91. doi:10.1007/s12021-013-9200-7

Demiris, Y., and Khadhouri, B. (2006). Hierarchical attentive multiple models for execution and recognition (HAMMER). *Rob. Auton. Syst.* 54, 361–369. doi:10.1016/j.robot.2006.02.003

Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., and Tanila, H. (1999). The hippocampus, memory, and place cells: is it spatial memory or a memory space? *Neuron* 23, 209–226. doi:10.1016/S0896-6273(00)80773-4

Flanagan, J. R., and Johansson, R. S. (2003). Action plans used in action observation. *Nature* 424, 769–771. doi:10.1038/nature01861

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119, 593–609. doi:10.1093/brain/119.2.593

Gallistel, C. R. (1989). Animal cognition: the representation of space, time and number. *Annu. Rev. Psychol.* 40, 155–189. doi:10.1146/annurev.ps.40.020189.001103

Gallistel, C. R. (1990). *The Organization of Learning*. Cambridge, MA: MIT Press.

Gaussier, P., Revel, A., Banquet, J. P., and Babeau, V. (2002). From view cells and place cells to cognitive map learning: processing stages of the hippocampal system. *Biol. Cybern.* 86, 15–28. doi:10.1007/s004220100269

Gilbert, D. T., and Wilson, T. D. (2007). Prospection: experiencing the future. *Science* 317, 1351–1354. doi:10.1126/science.1144161

Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behav. Brain Sci.* 27, 377–442. doi:10.1017/S0140525X04000093

Harnad, S. (1990). The symbol grounding problem. *Physica D* 42, 335–346. doi:10.1016/0167-2789(90)90087-6

Hesslow, G. (2002). Conscious thought as simulation of behaviour and perception. *Trends Cogn. Sci.* 6, 242–247. doi:10.1016/S1364-6613(02)01913-7

Hesslow, G. (2012). The current status of the simulation theory of cognition. *Brain Res.* 1428, 71–79. doi:10.1016/j.brainres.2011.06.026

Hommel, B., Müsseler, J., Aschersleben, G., and Prinz, W. (2001). The theory of event coding (TEC): a framework for perception and action planning. *Behav. Brain Sci.* 24, 849–937. doi:10.1017/S0140525X01000103

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annu. Rev. Psychol.* 60, 653–670. doi:10.1146/annurev.psych.60.110707.163604

Johansson, R. S., Westling, G., Bäckström, A., and Flanagan, J. R. (2001). Eye-hand coordination in object manipulation. *J. Neurosci.* 21, 6917–6932.

Kleinberg, J. (2006). "Complex networks and decentralized search algorithms," in *Proceedings of the International Congress of Mathematicians (ICM)*. Madrid, 1019–1044.

Krichmar, J. L., and Edelman, G. M. (2005). Brain-based devices for the study of nervous systems and the development of intelligent machines. *Artif. Life* 11, 63–77. doi:10.1162/1064546053278946

Krichmar, J. L., and Edelman, G. M. (2006). "Principles underlying the construction of brain-based devices," in *Proceedings of AISB '06 - Adaptation in Artificial and Biological Systems*, Volume 2 of *Symposium on Grand Challenge 5: Architecture of Brain and Mind*, eds T. Kovacs, and J. A. R. Marshall (Bristol: University of Bristol), 37–42.

Krüger, N., Geib, C., Piater, J., Petrickb, R., Steedman, M., Wörgötter, F., et al. (2011). Object–action complexes: grounded abstractions of sensory–motor processes. *Rob. Auton. Syst.* 59, 740–757. doi:10.1016/j.robot.2011.05.009

Lowe, R., and Ziemke, T. (2011). The feeling of action tendencies: on the emotional regulation of goal-directed behaviours. *Front. Psychol.* 2:1–24. doi:10.3389/fpsyg.2011.00346

Lukic, L., Santos-Victor, J., and Billard, A. (2012). " Learning coupled dynamical systems from human demonstration for robotic eye-arm-hand coordination," in *Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, (Osaka, Japan), 552–559.

McNamara, T. P., and Shelton, A. L. (2003). Cognitive maps and the hippocampus. *Trends Cogn. Sci.* 7, 333–335. doi:10.1016/S1364-6613(03)00167-0

McNaugton, B., Battaglia, F., Jensen, O., Moser, E., and Moser, M. (2006). Path integration and the neural basis of the 'cognitive map'. *Nat. Rev.* 7, 663–678. doi:10.1038/nrn1932

Merrick, K. E. (2010). A comparative study of value systems for self-motivated exploration and learning by robots. *IEEE Trans. Auto. Mental Dev.* 2, 119–131. doi:10.1109/TAMD.2010.2051435

Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., et al. (2010). The iCub humanoid robot: an open-systems platform for research in cognitive development. *Neural Netw.* 23, 1125–1134. doi:10.1016/j.neunet.2010.08.010

Mohan, V., Sandini, G., and Morasso, P. (2014). A neural framework for organization and flexible utilization of episodic memory in cumulatively learning baby humanoids. *Neural Comput.* 26, 1–43. doi:10.1162/NECO_a_00664

Morse, A., Lowe, R., and Ziemke, T. (2008). "Towards an enactive cognitive architecture," in *Proceedings of the First International Conference on Cognitive Systems*, (Karlsruhe, Germany).

Moulton, S. T., and Kosslyn, S. M. (2009). Imagining predictions: mental imagery as mental emulation. *Philos. Trans. R. Soc. B* 364, 1273–1280. doi:10.1098/rstb.2008.0314

Newman, M. E. J. (2000). Models of the small world: a review. *J. Stat. Phys.* 101, 819–841. doi:10.1023/A:1026485807148

O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Exp. Neurol.* 51, 78–109. doi:10.1016/0014-4886(76)90055-8

O'Keefe, J., and Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Oxford: Clarendon.

Ondobaka, S., and Bekkering, H. (2012). Hierarchy of idea-guided action and perception-guided movement. *Front. Cogn.* 3:1–5. doi:10.3389/fpsyg.2012.00579

Østby, Y., Walhovd, K. B., Tamnes, C. K., Grydeland, H., Westlye, L. G., and Fjell, A. M. (2012). Mental time traval and default-mode network functional connectivity in the developing brain. *PNAS* 109, 16800–16804. doi:10.1073/pnas.1210627109

Rizzolatti, G., and Craighero, L. (2004). The mirror neuron system. *Annu. Rev. Physiol.* 27, 169–192. doi:10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., and Fadiga, L. (1998). "Grasping objects and grasping action meanings: the dual role of monkey rostroventral premotor cortex (area F5)," in *Sensory Guidance of Movement, Novartis Foundation Symposium 218*, eds G. R. Bock, and J. A. Goode (Chichester: John Wiley and Sons), 81–103.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.* 3, 131–141. doi:10.1016/0926-6410(95)00038-0

Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson's University Library.

Sandini, G., Metta, G., and Vernon, D. (2007). "The iCub cognitive humanoid robot: An open-system research platform for enactive cognition," in *50 Years of AI*, Vol. 4850, eds M. Lungarella, F. Iida, J. C. Bongard, and R. Pfeifer (Heidelberg: Springer), 359–370.

Schacter, D. L., and Addis, D. R. (2007a). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philos. Trans. R. Soc. B* 362, 773–786. doi:10.1098/rstb.2007.2087

Schacter, D. L., and Addis, D. R. (2007b). Constructive memory – the ghosts of past and future: a memory that works by piecing together bits of the past may be better suited to simulating future events than one that is a store of perfect records. *Nature* 445, 27. doi:10.1038/445027a

Schacter, D. L., Addis, D. R., and Buckner, R. L. (2008). Episodic simulation of future events: concepts, data, and applications. *Ann. N. Y. Acad. Sci.* 1124, 39–60. doi:10.1196/annals.1440.001

Seligman, M. E. P., Railton, P., Baumeister, R. F., and Sripada, C. (2013). Navigating into the future or driven by the past. *Perspect. Psychol. Sci.* 8, 119–141. doi:10.1177/1745691612474317

Shanahan, M. P. (2005a). "Cognition, action selection, and inner rehearsal," in *Proceedings IJCAI Workshop on Modelling Natural Action Selection*, 92–99.

Shanahan, M. P. (2005b). "Emotion, and imagination: a brain-inspired architecture for cognitive robotics," in *Proceedings AISB 2005 Symposium on Next Generation Approaches to Machine Consciousness*, 26–35.

Shanahan, M. P. (2006). A cognitive architecture that combines internal simulation with a global workspace. *Conscious. Cogn.* 15, 433–449. doi:10.1016/j.concog.2005.11.005

Shanahan, M. P., and Baars, B. (2005). Applying global workspace theory to the frame problem. *Cognition* 98, 157–176. doi:10.1016/j.cognition.2004.11.007

Sloman, A. Available at: http://www.cs.bham.ac.uk/research/projects/cogaff/misc/talks/models.pdf.

Squire, L. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiol. Learn. Mem.* 82, 171–177. doi:10.1016/j.nlm.2004.06.005

Stachenfeld, K., Botvinick, M., and Gershman, S. J. (2014). "Design principles of the hippocampal cognitive map," in *Advances in Neural Information Processing Systems*. eds Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (New York: Curran Associates, Inc.), 2528–2536.

Sterling, P. (2004). "Principles of allostasis," in *Allostasis, Homeostasis, and the Costs of Adaptation*, ed. J. Schulkin (Cambridge: Cambridge University Press), 17–64.

Sterling, P. (2012). Allostasis: a model of predictive regulation. *Physiol. Behav.* 106, 5–15. doi:10.1016/j.physbeh.2011.06.004

Stock, A., and Stock, C. (2004). A short history of ideo-motor action. *Psychol. Res.* 68, 176–188. doi:10.1007/s00426-003-0154-5

Svensson, H., Lindblom, J., and Ziemke, T. (2007). "Making sense of embodied cognition: simulation theories of shared neural mechanisms for sensorimotor and cognitive processes," in *Body, Language and Mind*, Vol. 1, eds T. Ziemke, J. Zlatev, and R. M. Frank (Berlin: Mouton de Gruyter), 241–269.

Svensson, H., Morse, A. F., and Ziemke, T. (2009). "Representation as internal simulation: a minimalistic robotic model," in *Proceedings of the Thirty-first Annual Conference of the Cognitive Science Society*, eds N. Taatgen, and H. van Rijn (Austin, TX: Cognitive Science Society), 2890–2895.

Svensson, H., Thill, S., and Ziemke, T. (2013). Dreaming of electric sheep? Exploring the functions of dream-like mechanisms in the development of mental imagery simulations. *Adapt. Behav.* 21, 222–238. doi:10.1177/1059712313491295

Swain, M., and Ballard, D. (1991). Color indexing. *Int. J. Comput. Vis.* 7, 11–32. doi:10.1007/BF00130487

Swain, M. J., and Ballard, D. H. (1990). "Indexing via colour histograms," in *International Conference on Computer Vision – ICCV90*, 390–393.

Szpunar, K. K. (2010). Episodic future throught: an emerging concept. *Perspect. Psychol. Sci.* 5, 142–162. doi:10.1177/1745691610362350

Telesford, Q. K., Joyce, K. E., Hayasaka, S., Burdette, J. H., and Laurienti, P. J. (2011). The ubiquity of small-world networks. *Brain Connect.* 1, 367–375. doi:10.1089/brain.2011.0038

Tenorth, M., and Beetz, M. (2009). "KnowRob – knowledge processing for autonomous personal robots," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 4261–4266.

Tenorth, M., Perzylo, A. C., Lafrenz, R., and Beetz, M. (2012). "The roboearth language: representing and exchanging knowledge about actions, objects, and environments," in *IEEE International Conference on Robotics and Automation*, 1284–1289.

Tenorth, M., Perzylo, A. C., Lafrenz, R., and Beetz, M. (2013). Representation and exchange of knowledge about actions, objects, and environments in the roboearth framework. *IEEE Trans. Auto. Sci. Eng.* 10, 643–651. doi:10.1109/TASE.2013.2244883

Thill, S., Caligiore, D., Borghi, A. M., Ziemke, T., and Baldassarre, G. (2013). Theories and computational models of affordance and mirror systems: an integrative review. *Neurosci. Biobehav. Rev.* 37, 491–521. doi:10.1016/j.neubiorev.2013.01.012

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychol. Rev.* 55, 189–208. doi:10.1037/h0061626

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–735. doi:10.1017/S0140525X05000129

Traver, V. J., and Bernardino, A. (2010). A review of log-polar imaging for visual perception in robotics. *Rob. Auton. Syst.* 58, 378–398. doi:10.1016/j.robot.2009.10.002

Tsagarakis, N. G., Metta, G., Sandini, G., Vernon, D., Beira, R., Santos-Victor, J., et al. (2007). iCub – the design and realisation of an open humanoid platform for cognitive and neuroscience research. *Int. J. Adv. Robot.* 21, 1151–1175. doi:10.1163/156855307781389419

Tsotsos, J. K. (2006). "Cognitive vision need attention to link sensing with recognition," in *Cognitive Vision Systems: Sampling the Spectrum of Approaches*, Vol. 3948, eds H. I. Christensen, and H.-H. Nagel (Heidelberg: Springer), 25–36.

Tsotsos, J. K. (2011). *A Computational Perspective on Visual Attention*. Cambridge, MA: MIT Press.

Tsotsos, J. K., Culhane, S., Wai, W., Lai, Y., David, N., and Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artif. Intell.* 78, 507–547. doi:10.1016/0004-3702(95)00025-9

Tulving, E. (1972). "Episodic and semantic memory," in *Organization of Memory*, eds E. Tulving, and W. Donaldson (New York: Academic Press), 381–403.

Tulving, E. (1983). *Elements of Episodic Memory*. Oxford: Oxford University Press.

Tulving, E. (1984). Précis of elements of episodic memory. *Behav. Brain Sci.* 7, 223–268. doi:10.1017/S0140525X0004440X

Ulanowicz, R. E. (2000). *Growth and Development; Ecosystems Phenomenology*. Lincoln: Excel Press.

Vernon, D. (2014). *Artificial Cognitive Systems – A Primer*. Cambridge, MA: MIT Press.

Vernon, D., Metta, G., and Sandini, G. (2007). A survey of artificial cognitive systems: implications for the autonomous development of mental capabilities in computational agents. *IEEE Trans. Evol. Comput.* 11, 151–180. doi:10.1109/TEVC.2006.890274

Vernon, D., von Hofsten, D., and Fadiga, L. (2010). *A Roadmap for Cognitive Development in Humanoid Robots, Cognitive Systems Monographs (COSMOS)*, Vol. 11, (Berlin: Springer).

von Hofsten, C. (2004). An action perspective on motor development. *Trends Cogn. Sci.* 8, 266–272. doi:10.1016/j.tics.2004.04.002

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of small world networks. *Nature* 393, 440–442. doi:10.1038/30918

Wintermute, S. (2012). Imagery in cognitive architecture: representation and control at multiple levels of abstraction. *Cogn. Syst. Res.* 1, 1–29. doi:10.1016/j.cogsys.2012.02.001

Wood, R., Baxter, P., and Belpaeme, T. (2012). A review of long-term memory in natural and synthetic systems. *Adapt. Behav.* 20, 81–103. doi:10.1177/1059712311421219

Zaharescu, A., Rothenstein, A. L., and Tsotsos, J. K. (2004). "Towards a biologically plausible active visual search model," in *Proceedings of the Second International Workshop on Attention and Performance in Computational Vision, WAPCV*, Vol. 3368, eds L. Paletta, J. K. Tsotsos, E. Rome, and G. Humphreys (Berlin: Springer), 133–147.

Ziemke, T., and Lowe, R. (2009). On the role of emotion in embodied cognitive architectures: from organisms to robots. *Cogn. Comput.* 1, 104–117. doi:10.1007/s12559-009-9012-0

frontiers
in Robotics and AI

Space as an Invention of Active Agents

Alexander V. Terekhov* and J. Kevin O'Regan

Laboratoire Psychologie de Perception, CNRS, University Paris Descartes, Paris, France

The question of the nature of space around us has occupied thinkers since the dawn of humanity, with scientists and philosophers today implicitly assuming that space is something that exists objectively. Here, we show that this does not have to be the case: the notion of space could emerge when biological organisms seek an economic representation of their sensorimotor flow. The emergence of spatial notions does not necessitate the existence of real physical space, but only requires the presence of sensorimotor invariants called "compensable" sensory changes. We show mathematically and then in simulations that naive agents making no assumptions about the existence of space are able to learn these invariants and to build the abstract notion that physicists call rigid displacement, independent of what is being displaced. Rigid displacements may underly perception of space as an unchanging medium within which objects are described by their relative positions. Our findings suggest that the question of the nature of space, currently exclusive to philosophy and physics, should also be addressed from the standpoint of neuroscience and artificial intelligence.

Keywords: sensorimotor contingencies, space perception, naive agent, concepts development, compensable transformation, geometry, artificial intelligence and robotics

1. INTRODUCTION

How do we know that there is space around us? Our brains sit inside the dark bony cavities formed by the skull, with myriads of sensorimotor signals coming in and going out. From this immense flow of spikes, our brains conclude that there is such thing as space, filled with such things as objects, and that there is such thing as body – a special type of object which brains have most control over. Taking this "tabula rasa" approach, it is not clear what constitutes space as something discoverable in the sensory information, or, in other words, how space manifests itself to a naive agent that has no information other than its undifferentiated sensory inputs and motor outputs.

Poincaré (1905) was among the first to recognize this problem and to attempt its mathematical formalization. He suggested that space can manifest itself through what he called "compensable changes": such changes in the world, which the agent can nullify by its own action. For example, consider standing in front of a red ball. The light reflected from the ball is projected into the retina where it creates excitation of the sensory cells. If now the ball displaces 1 m away the input to the retina becomes different from what it was before. Yet, you can make the input to be the same as before if you walk 1 m in the same direction as the ball. It is through this ability to nullify the changes in the environment that we learn about space (Poincaré, 1905). This approach was further developed by Nicod (1929), who showed, among other things, that temporal sequences can be used to determine the topology of space. After Nicod, this line of research was for long time discontinued, until it was reinitiated in the field of artificial intelligence and robotics (Kuipers, 1978; Pierce and Kuipers, 1997). Nowadays, a whole body of work has accumulated describing how robotic agents can build

models of themselves and their environments (Kaplan and Oudeyer, 2004; Klyubin et al., 2004, 2005; Gloye et al., 2005; Bongard et al., 2006; Hersch et al., 2008; Hoffmann et al., 2010; Gordon and Ahissar, 2011; Sigaud et al., 2011; Koos et al., 2013). However, the question of the acquisition of the spatial concepts as something independent of particular sensory coding remains rather poorly studied [however, see Philipona et al. (2003), Roschin et al. (2011), and Laflaquiere et al. (2012)].

In the current paper, we show how a naive agent can acquire spatial notions in the form of internalized (or "sensible," cf., Nicod, 1929) rigid displacements. We show that being equipped with such notions the agent can solve spatial tasks that would be unsolvable in the metric of the original sensory inputs. Moreover, we show that notions indistinguishable from internalized rigid displacements can be built by an agent inhabiting a spaceless universe. We thus suggest that the notion of space we possess is a construct of our perceptual system, based on certain sensorimotor invariants, which, however, do not necessitate the objective existence of space.[1]

2. ILLUSTRATION OF PRINCIPLE

To illustrate the principle, consider first the sensory universe or "Merkwelt" (cf von Uexküll, 1957) of the one-dimensional agent in **Figure 1**. Note that in the present work, we are attempting a proof of concept showing that an agent interacting with the world could adduce the notion of space. For this reason, we will be assuming that the agent is equipped with sufficient memory and
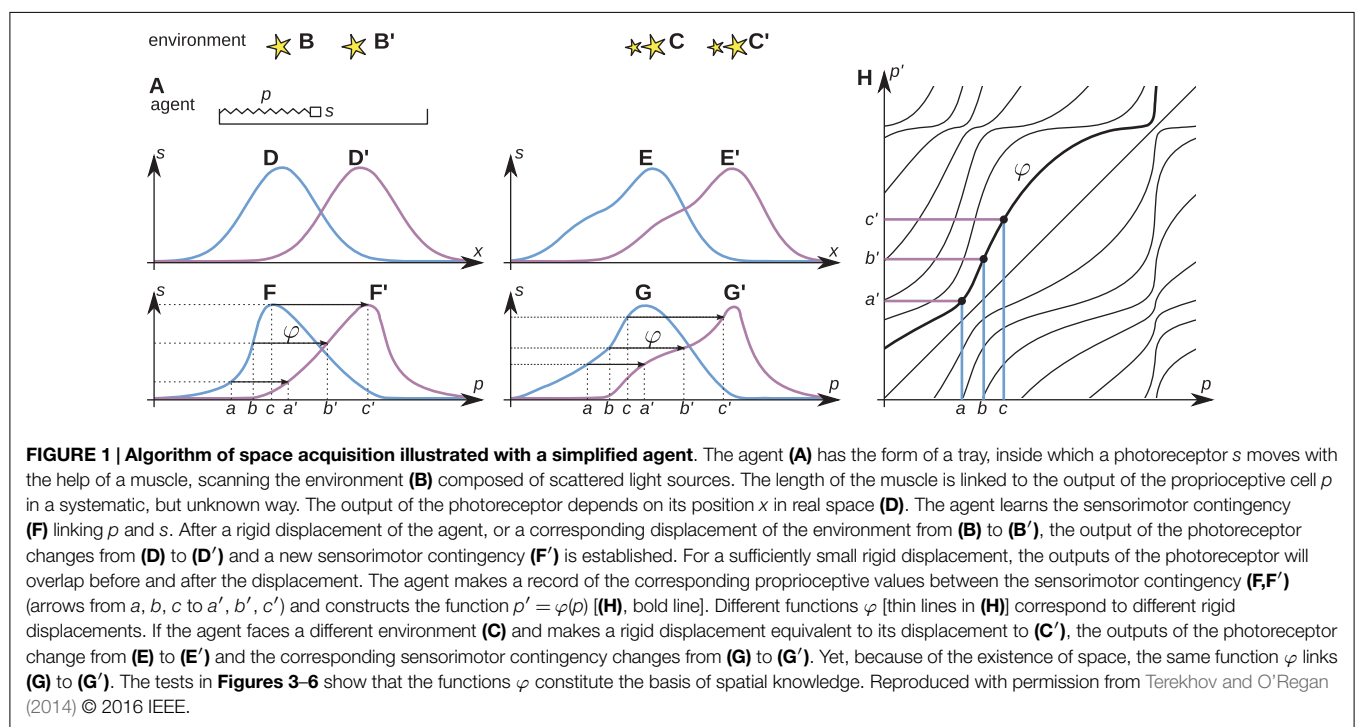
---

[1]While the present article concerns the notion of space, it would be extremely interesting to attempt a similar approach for the emergence of the notion of time. However at present we have no clear idea of how to do this. In the present article we have attempted to reduce assumptions about time to a minimum.

computational resources to perform the necessary manipulation of the sensorimotor information.

Assume (though this is not known to the agent's brain) that its body is composed of a single photoreceptive sensor that can move laterally inside its body using a "muscle" (**Figure 1A**). Assume a one-dimensional environment as in **Figure 1B**, and assume first that it is static. If the agent were to perform scanning actions with the muscle and were to plot photoreceptor output against the photoreceptor's actual physical position, it would obtain a plot such as **Figure 1D**. But it cannot do this because it has no notion, let alone any measure, of physical position, and only has knowledge of proprioception. The agent can only plot photoreceptor output against proprioception, and so obtains a distorted plot as in **Figure 1F**. This "sensorimotor contingency" (MacKay, 1962; O'Regan and Noë, 2001) is all that the agent knows about. It does not know anything about the structure of its body and sensor, let alone that there is such a thing as space in which it is immersed. Indeed, the agent does not need such notions to understand its world, since its world is completely accounted for by its knowledge of the sensorimotor contingency it has established by scanning.

But now suppose that the environment can move relative to the agent, for example, taking **Figures 1B,B′**. The previously plotted sensorimotor contingency will no longer apply, and a different plot will be obtained (e.g., **Figure 1F′**). The agent goes from being able to completely predict the effects of its scanning actions on its sensory input, to no longer being able to do so.

However, there is a notable fact which applies. Although the agent does not know this, physicists looking from outside the agent would note that if the displacement relative to the environment is not too large, there will be some overlap between the physical locations scanned before and after the displacement. In this overlapping region, the sensor occupies the same

FIGURE 1 | Algorithm of space acquisition illustrated with a simplified agent. The agent **(A)** has the form of a tray, inside which a photoreceptor s moves with the help of a muscle, scanning the environment **(B)** composed of scattered light sources. The length of the muscle is linked to the output of the proprioceptive cell p in a systematic, but unknown way. The output of the photoreceptor depends on its position x in real space **(D)**. The agent learns the sensorimotor contingency **(F)** linking p and s. After a rigid displacement of the agent, or a corresponding displacement of the environment from **(B)** to **(B′)**, the output of the photoreceptor changes from **(D)** to **(D′)** and a new sensorimotor contingency **(F′)** is established. For a sufficiently small rigid displacement, the outputs of the photoreceptor will overlap before and after the displacement. The agent makes a record of the corresponding proprioceptive values between the sensorimotor contingency **(F,F′)** (arrows from a, b, c to $a′$, $b′$, $c′$) and constructs the function $p′ = \varphi(p)$ [**(H)**, bold line]. Different functions φ [thin lines in **(H)**] correspond to different rigid displacements. If the agent faces a different environment **(C)** and makes a rigid displacement equivalent to its displacement to **(C′)**, the outputs of the photoreceptor change from **(E)** to **(E′)** and the corresponding sensorimotor contingency changes from **(G)** to **(G′)**. Yet, because of the existence of space, the same function φ links **(G)** to **(G′)**. The tests in **Figures 3–6** show that the functions φ constitute the basis of spatial knowledge. Reproduced with permission from Terekhov and O'Regan (2014) © 2016 IEEE.

positions relative to the environment as it occupied before the displacement occurred. Since sensory input depends only on the position of the photoreceptor relative to the environment, the agent will thus discover that for these positions the sensory input from the photoreceptor will be the same as before the displacement.

Registering such a coincidence is not uncommon for an agent with a single photoreceptor, but the same would happen for an agent with numerous receptors. For such a more complicated agent, the coincidence would be extremely noteworthy.

In an attempt to better "understand" its environment, the agent will thus naturally make a catalog of these coincidences (cf. arrows in **Figures 1F,F′**), and so establish a function φ linking the values of proprioception observed before a change to the corresponding values of proprioception after the change. Such a function for all values of proprioception is shown in **Figure 1H**.

Assume that over time, the environment displaces rigidly to various extents, with the agent located initially at various positions. Furthermore, assume that such displacements can happen for entirely different environments (e.g., **Figure 1C**). Since the sensorimotor contingencies themselves depend on all these factors, it might be expected that different functions φ would have to be cataloged for all these different cases. Yet, it is a remarkable fact that the set of functions φ is much simpler: for a given displacement of the environment, the agent will discover the *same* functions φ, even when this displacement starts from different initial positions, and even when the environment is different.

We shall see below that this remarkable simplicity of the functions φ *provides the agent with the notion of space*. But, first let us see where the simplicity derives from.

Each function φ links proprioceptive values before an environmental change to proprioceptive values after the change, in such a way that for the 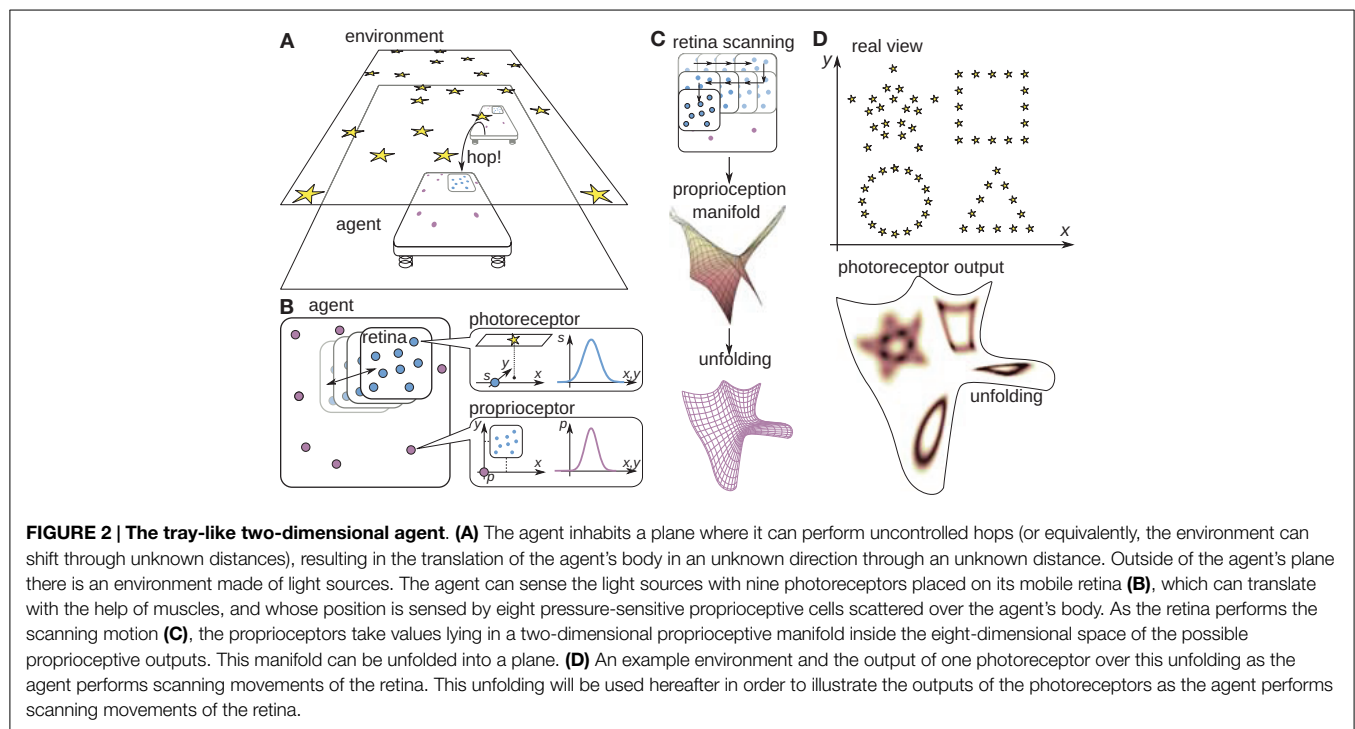linked values the outputs of the photoreceptor match before and after the change. Seen from outside the agent, the physicist would know that this situation will occur if the agent's photoreceptor occupies the same position relative to the environment before and after the environmental change. And this will happen *if (1) the environment makes a rigid displacement, and if (2) the agent's photoreceptor makes a rigid displacement equal to the rigid displacement of the environment*. Thus, physicists looking at the agent would know that the functions φ actually measure, in proprioceptive coordinates, rigid physical displacements of the environment relative to the agent (or vice versa) (see Section 6).

Let us stress again that *a priori* there was no reason at all why the φ functions for different starting points should be the same for a given displacement, and the same for all environments. But now, we can understand why the set of functions φ is so simple: it is because a defining property of rigid displacements is that they are independent of their starting points, and independent of the properties of what is being displaced.

The functions φ can thus be seen as perceptual constructs equivalent to physical rigid displacements, or one could say, following Nicod (1929), that they are *sensible rigid displacements*, where *sensible* refers to the fact that they are *defined within the Merkwelt of the agent*.

3. RESULTS

To illustrate that these sensible rigid displacements or functions φ really have the properties of real physical rigid displacements, we will use computer simulations with the more complex two-dimensional agent described in **Figure 2**. The details of the simulation are presented in the Methods (section 5). In Formalization (section 6), we show that the demonstration applies to an arbitrarily complicated agent, with certain restrictions.

FIGURE 2 | The tray-like two-dimensional agent. **(A)** The agent inhabits a plane where it can perform uncontrolled hops (or equivalently, the environment can shift through unknown distances), resulting in the translation of the agent's body in an unknown direction through an unknown distance. Outside of the agent's plane there is an environment made of light sources. The agent can sense the light sources with nine photoreceptors placed on its mobile retina **(B)**, which can translate with the help of muscles, and whose position is sensed by eight pressure-sensitive proprioceptive cells scattered over the agent's body. As the retina performs the scanning motion **(C)**, the proprioceptors take values lying in a two-dimensional proprioceptive manifold inside the eight-dimensional space of the possible proprioceptive outputs. This manifold can be unfolded into a plane. **(D)** An example environment and the output of one photoreceptor over this unfolding as the agent performs scanning movements of the retina. This unfolding will be used hereafter in order to illustrate the outputs of the photoreceptors as the agent performs scanning movements of the retina.

In **Figure 3**, the two-dimensional simulated agent is first shown an environment that makes a certain displacement (or the agent makes an equivalent hop relative to the environment). The agent is then shown two other instances of the same displacement, but with two completely different environments. Even though in each case, the sensory experiences of the agent are different, and even though they change in different ways, the Figure shows that *the same function* φ can be used to account for these changes. This is what is to be expected from a notion of rigid displacement, which should not depend on the content of what is displaced.

Figure 4 shows further that, once equipped with the notion of sensible rigid displacement, the agent is well on its way toward understanding space. In particular, sensible rigid displacements endow the agent with the percept of space as an *unchanging* medium, which implies being able to distinguish between the sensory changes caused by the proper movements of the agent from those reflecting the deformation of the environment. **Figure 4** shows how the simulated agent is able to distinguish between the two despite the fact that in the sensory input a rigid shift may look like a deformation (**Figure 4C**), while deformations may seem just like a minor displacement (**Figure 4B**).

Figure 5 shows that the agent can define the notion of *relative position* of A with respect to B. This notion is more abstract than displacement, as there exist numerous paths leading from B to A, while relative position is independent of the choice of a particular path. The notion of relative position allows the agent to "understand" that it is at the same "somewhere" independently of how it got there. To define the notion of relative position, the agent must be able to take different combinations of displacements having the same origin and destination, and consider them as equivalent.

4. DISCUSSION

We have shown that, without assuming *a priori* the existence of space, the agent invents the notions of *sensible displacement*, *unchanging medium* and *relative position*. These notions allow the agent to conceive of its environment in a way that we can assimilate to possessing the notion of space. The agent can now separate the properties of its sensed environment into properties a physicist would call *spatial* (position, orientation) and *non-spatial* (shape, color, etc). These are the properties whose changes the agent respectively can and cannot account for in terms of sensible rigid displacements. Several further points should be mentioned.

The method that the agent uses to "invent" its notion of space involves defining φ functions from matching sensory signals. As is the case for temporal coincidence, this can be understood as a strategy of associating causes that lead to the same consequences (Markram et al., 2011). This is a productive learning strategy in general and is easily implementable in neural hardware.

Note, however, that constructing sensible rigid displacements on the basis of matches is only possible if sensory changes caused by modifications in the environment can be compensated (i.e., equalized or canceled) by the agent's own action. The conditions

for this to be possible are (1) that the agent be able to act,[2] and (2) that appropriate compensatory changes can occur in the environment. The agent's own actions are thus crucial for the acquisition of the notion of space. Of course, if the agent knows in advance that there is space, it may be able to reconstruct it without acting. But if the agent has limited action capacities, it will not invent space "correctly." In particular, the simple two-dimensional agent we have considered has a retina that can translate, but cannot rotate. This agent will therefore classify relative position, but not orientation, as being a spatial property. Evidence from biology also shows the importance of action in the acquisition of spatial notions: an example is the classic result of Held and Hein (1963).

In addition to action, sufficient richness of the environment is essential for an agent to discover space. If for example displacement in a certain direction has no sensory consequences, or if they are ambiguous, then the agent will be unable to learn the corresponding sensible rigid displacements. Again this is coherent with biology, where it has been shown (Blakemore and Cooper, 1970) that kittens raised in visual environments composed of vertical stripes are blind to displacements of horizontally aligned objects and vice versa.

Another point worth mentioning is the fact that sensible rigid displacements are nothing but abstract constructs – they do not imply that something really moves: if the agent inhabited a different physical universe but where the sensorimotor regularities were the same, then it would develop the same construct of sensible rigid displacement. For example in Audio Agent (section 7), we describe an agent whose world consists only of sounds, but that develops sensible rigid displacements in pitch analogous to the spatial constructs of the agent in **Figure 1**.

A final point concerns the statistical approaches often used up to now to understand brain functioning (Zhaoping, 2006; Ganguli and Sompolinsky, 2012). Such approaches use statistical correlations to compress the data observed in sensory and motor activity. It is possible that these approaches may be adapted to capture the "algebraic" notion of mutual compensability between environment changes and an agent's actions that is instantiated by the functions φ and that is essential for understanding the essence of space.

In conclusion, the three-dimensional space we perceive could be nothing but a construct, which simplifies the representation of information provided by our limited senses in response to our limited actions. In reality space – if it exists – may have a higher number of dimensions, most of which we perceive as non-spatial properties because of our inability to perform corresponding compensatory movements. Or, conversely, there may in fact be no physical space: our impression that space exists may be nothing but a gross oversimplification generated by our

---

[2]In the particular case of our agent the action involves moving the sensor within the agent's body. This ensures that the agent has a reliable measure of the motion that it is producing, with, in particular, a one-to-one relation between muscle changes and physical changes. Our algorithm would have to be improved in order to allow cases where the agent moved its body using, for example, legs whose repeated action creates motion, since here there is no longer a one-to-one link between leg muscle command and physical change in space.

perceptual systems, with the real world only being very approximately describable as a collection of "objects" moving through an "unchanging medium."

5. METHODS

5.1. Agent

The two-dimensional agent from **Figure 2** was simulated to illustrate the acquisition of spatial knowledge. The agent has a square body in a form of a tray, within which a square retina translates. We choose the measurement units so that the retina movements are confined to a unit square. The position x, y of the retina is registered by proprioceptors scattered over the body surface and having outputs

$$p_j = \exp\left\{-\frac{(d_j^p)^2}{(\sigma_j^p)^2}\right\},$$

where d_j^p is the distance between the center of the retina and the location of the j-th proprioceptor, and σ_j^p is its acuity.

The retina is covered with photoreceptors, measuring the intensity of the light coming from N_ℓ spot light sources located in a plane above the agent. The response of j-th photoreceptor is

$$s_j = \sum_{i=1}^{N_\ell} I_i \exp\left\{-\frac{(d_{ij}^s)^2}{(\sigma_j^s)^2}\right\},$$

where d_{ij}^s is the distance between the projection of the i-th light source onto the plane of the agent and the j-th photoreceptor; I_i is the intensity of the i-th light source, and σ_j^s is the acuity of the j-th photoreceptor.
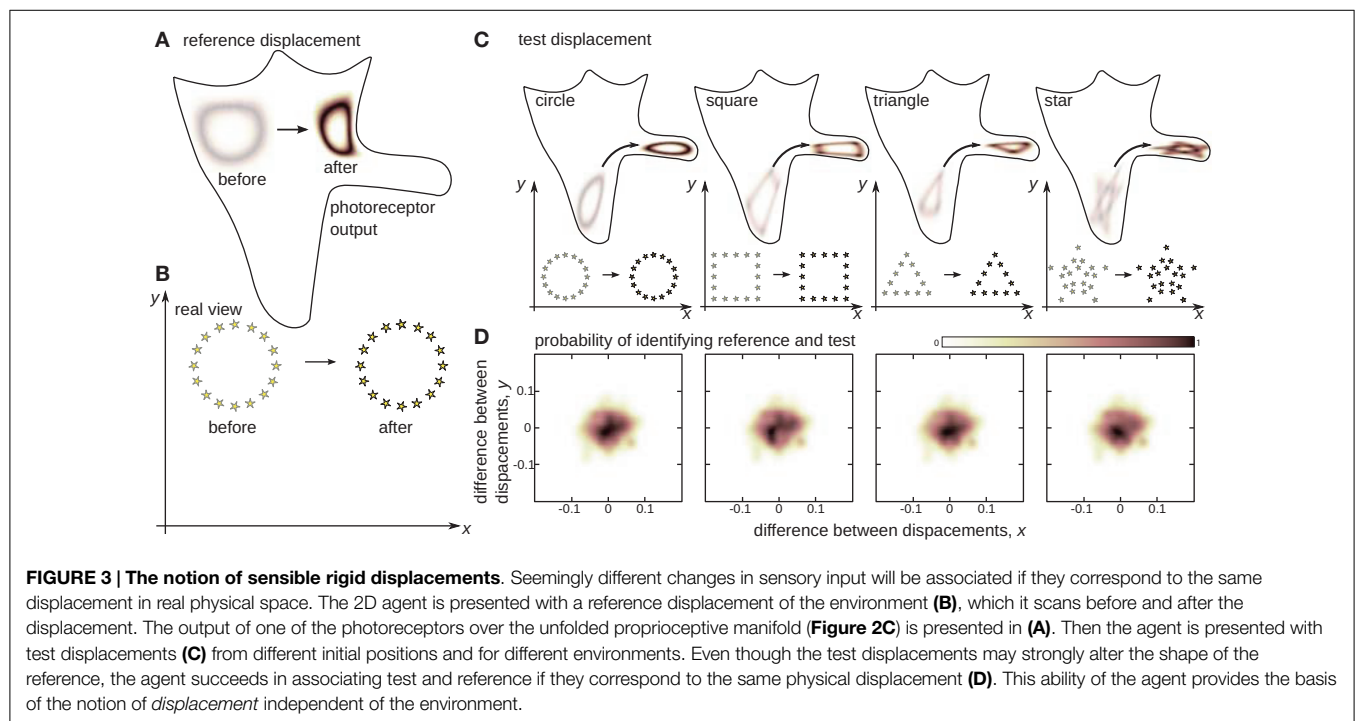
For the simulations presented in the paper we deliberately distributed the eight proprioceptors over the agent's body in a non-uniform way so as to ensure a certain amount of distortion of the image in **Figures 3–5**. Their acuities σ_j^p was set to 0.3 for all receptors. The positions of the nine photoreceptors were drawn randomly from a square with sides of length 0.3. The acuity of the receptors σ_j^s took random values between 0.03 and 0.3. Due to the retinal mobility the agent's "field of view" was a 1.0×1.0 square centered at what we call the agent's position.

5.2. Learning Functions φ

The agent was placed into the environment with 200 light sources distributed randomly in 3×3 square, centered at the agent's initial position (see **Figure 6**). The agent scanned the environment by moving the retina inside the body and tabulating the tuples of proprioceptive and photoreceptive inputs $\langle p_k, s_k \rangle$. The agent then jumped to a new position, which was within a 1.8×1.8 square centered at its initial position, and again scanned the environment and tabulated the tuples $\langle p_k', s_k' \rangle$. The agent then looked for the co-occurrences $s_k = s_{k'}'$ and put the corresponding proprioceptive inputs into pairs $\langle p_k, p_{k'}' \rangle$. The function φ was then defined as the set of all such pairs.

Exclusively for the sake of code optimization when "scanning" the environment the retina moved over a regular 201×201 grid. The outputs of the photoreceptors were considered as matching if for every photoreceptor the difference of the outputs before and after the jump was less than 0.005. The corresponding values of proprioception before and after the jump were taken to form the function φ. If the value of every proprioceptor in one pair differed by less than 0.01 from the value of proprioceptor in the other pair, then one of the pairs was discarded. The destination points of the agent's jumps also belonged to a regular grid centered at the agent's initial position and having a step size of 0.02. In total, we obtained 8281 different functions φ.

FIGURE 3 | The notion of sensible rigid displacements. Seemingly different changes in sensory input will be associated if they correspond to the same displacement in real physical space. The 2D agent is presented with a reference displacement of the environment **(B)**, which it scans before and after the displacement. The output of one of the photoreceptors over the unfolded proprioceptive manifold (**Figure 2C**) is presented in **(A)**. Then the agent is presented with test displacements **(C)** from different initial positions and for different environments. Even though the test displacements may strongly alter the shape of the reference, the agent succeeds in associating test and reference if they correspond to the same physical displacement **(D)**. This ability of the agent provides the basis of the notion of *displacement* independent of the environment.
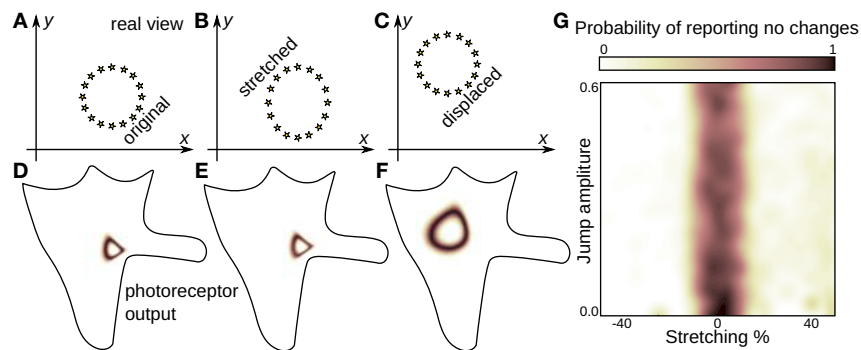
FIGURE 4 | The notion of space as unchanging medium. The 2D agent can distinguish between the sensory changes provoked by its own movement and those resulting from the joint effect of its own motion and changes in the environment. The agent is presented with an environment **(A)** which it scans **(D)**. Then the agent makes a jump and simultaneously the environment is stretched or shrunk along one axis by a certain amount **(B)**, which can be zero **(C)**. The agent is to judge whether the environment was the same before and after the jump. Note that the visual input in the modified environment **(E)** resembles the original **(D)** more than it does the unchanged environment **(F)**. Yet, the agent can successfully identify the case when the environment does not change, and it can do this independently of the extent of the jumps **(G)**. This ability of the agent underlies the notion of space as an *unchanging medium* through which the agent makes *displacements*.
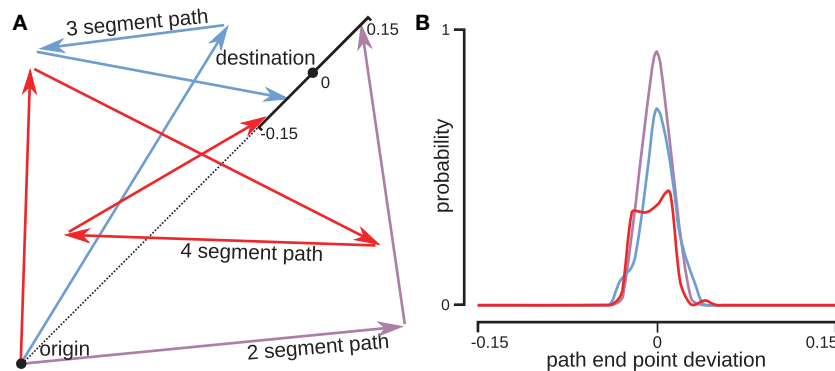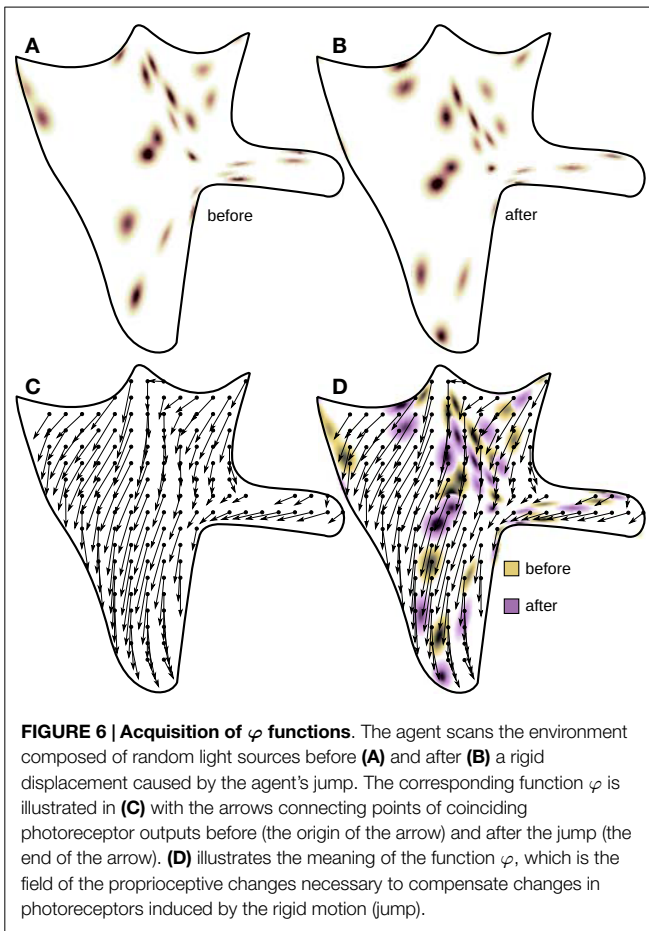

FIGURE 5 | The notion of relative position, independent of the connecting path. The 2D agent can construct the notion of *relative position* of a destination point with respect to an origin by associating all possible paths connecting the origin to the destination. The agent was presented with a reference single-segment path. Then it was returned to the origin and moved along two-, three-, or four-segment paths with possibly different destination points at a distance in the range [−0.15, 0.15] from the destination of the reference path; all points were chosen on the same straight line **(A)**. The surrounding environment was altered at every path presentation in order to ensure that the agent calculated the displacement instead of comparing the view at the destination points. The agent successfully associated together paths which arrived close to the original destination point. This association was more accurate for the 2-segment path, and became weaker as the number of path segments increased **(B)**. This can be explained by accumulation of the integration error.

It must be emphasized here that though we used only rigid displacements of the environment for learning the φ functions, the result would have been essentially the same if arbitrary deformations of the environment were allowed. For instance, in our pilot simulations, we allowed the environment to shift and then deform along one of the axes, and then computed the corresponding φ functions. We found that the φ functions for such non-rigid changes of the environment contained less than 3% of what a pure rigid displacement would contain and depended heavily upon the particular environment used. Thus, introducing a simple criterion, like retaining only those φ functions with a certain number of points, and running the simulations for both rigid and non-rigid changes would produce essentially the same functions φ as running the simulations for rigid displacements only.

5.3. Sensible Rigid Displacement

The agent was facing 40 light sources distributed uniformly along a circle with 0.1 radius. The center of the circle was chosen randomly within a 1.0×1.0 square centered at the agent. In the reference displacement, all stars moved as a whole to a new random position, which was also within a 1.0×1.0 square. The agent determined the function φ corresponding to the reference displacement. Then the agent was shown one of four objects shown in **Figure 3**: the same circle, a square (composed of 40 lights), a triangle (39 lights), or a star (40 lights). The square and triangle had sides of length 0.2, and the star had a ray length 0.3. The objects underwent a random test displacement with initial and final positions within a 1.0×1.0 square. In order to save simulation time, we only considered displacements which differed from the reference by no more than 0.1 for each axis. The agent

FIGURE 6 | Acquisition of φ functions. The agent scans the environment composed of random light sources before **(A)** and after **(B)** a rigid displacement caused by the agent's jump. The corresponding function φ is illustrated in **(C)** with the arrows connecting points of coinciding photoreceptor outputs before (the origin of the arrow) and after the jump (the end of the arrow). **(D)** illustrates the meaning of the function φ, which is the field of the proprioceptive changes necessary to compensate changes in photoreceptors induced by the rigid motion (jump).

determined the functions $\widetilde{\varphi}$ for each of the tests and computed the distance between φ and $\widetilde{\varphi}$ as

$$\rho(\varphi, \widetilde{\varphi}) = \sum_{k',\tilde{k}':p_k=\tilde{p}_{\tilde{k}}} \|p'_{k'} - \tilde{p}'_{\tilde{k}'}\|, \tag{1}$$

where $\|\cdot\|$ is a euclidean distance in the proprioception space. The agent identified two displacements as the same if the error was below a threshold which was chosen so that 90% of displacements of size less than 0.005 were considered identical. The procedure was repeated 1,000 times.

5.4. Unchanging Medium
The agent was facing 40 light sources distributed uniformly over a circle with radius 0.1. The center of the circle was chosen randomly in a 0.4×0.4 square centered at the agent. The agent scanned the environment and tabulated the tuples $\langle p_k, s_k \rangle$. Then the agent made a jump to a random point located in a 0.6×0.6 square and simultaneously the circle was randomly stretched or shrunk by up to 50% along a fixed axis. Only those jump destinations were considered for which the agent could "see" the entire circle. The agent scanned the environment again and tabulated new tuples $\langle p'_k, s'_k \rangle$. The agent then searched for a function φ which gave the best fit of the photoreceptors after the jump based on

their values before the jump. In particular, the following error was computed:

$$\varepsilon = \sum_{k=1}^{N_k} \|s_k - s'_{k'}\|,$$

where k' was such that $\varphi(p_k) = p'_{k'}$. If the error was below the threshold, the agent assumed that the environment did not change during the jump. The threshold value of the error was chosen in such a way that the agent answered correctly in 90% of cases when the deformation of the circle was below 0.5%. **Figure 2** shows the result of simulations computed on the basis of 10,000 repetitions of the test.

5.5. Relative Position
The agent was facing an environment filled with 200 light sources with random locations and intensities. It was displaced from its original position to the destination point, which had coordinates (0.6, 0.6) relative to the agent's initial position. The agent determined the reference function φ_{ref}, which gave the best account of the displacement-induced changes of the photoreceptor outputs. Then the environment was replaced with a new randomly generated environment, and the agent was moved along a path composed of several segments. At every intermediate point along the path, the agent determined the function φ_j accounting for the changes in photoreceptor values. The agent then computed the composition function $\varphi_{\mathrm{comp}} = \varphi_n \circ \cdots \circ \varphi_1$, where n is the number of path segments. For any two functions φ and $\widetilde{\varphi}$ defined by sets of pairs $\langle p_k, p'_{k'} \rangle$ and $\langle \tilde{p}_{\tilde{k}}, \tilde{p}'_{\tilde{k}'} \rangle$ the composition $\widetilde{\varphi} \circ \varphi$ was defined as a set of pairs $\langle p_k, \tilde{p}'_{\tilde{k}'} \rangle$, such that $p'_{k'} = \tilde{p}_{\tilde{k}}$. The distance between φ_{ref} and φ_{comp} was computed using formula 1. The test and reference paths were assumed to correspond to the same relative position if the distance was below the same threshold as for the sensible rigid displacements. The procedure was repeated 1,000 times for two-, three-, and four-segment paths. Each intermediate point of the path was within the 0.9×0.9 square centered at the original position. In order to reduce simulation time the final points of all paths lay on the same line and were not more than 0.1 away from the origin.

6. FORMALIZATION
Here, we consider a general agent immersed in real physical space. Later, we will abandon the assumption of the existence of physcial space and give the conditions for the emergence of perceptual "space-like" constructs independently of whether they correspond to any real physical space.

Let s be the vector of the agent's exteroceptor outputs. The exteroceptors are connected to a body, assumed to be rigid, whose position and orientation is described by a spatial coordinate defined by the vector x. For every environment $\mathcal{E}$, the outputs of the exteroceptors are defined by a function

$$s = \sigma_{\mathcal{E}}(x). \tag{2}$$

We assume that this function has the property that if the environment $\mathcal{E}$ makes a rigid motion and becomes $\mathcal{E}'$, then there exists a rigid transformation T of entire space such that

$$s' = \sigma_{\mathcal{E}'}(x) = \sigma_{\mathcal{E}}(T(x)). \tag{3}$$

Proprioception p reports the position of the exteroceptors in the agent's body. For a given position of the agent $\mathcal{X}$ we assume there is a function $\pi_{\mathcal{X}}$ such that

$$p = \pi_{\mathcal{X}}(x). \quad (4)$$

Again, the function $\pi_{\mathcal{X}}$ has the property that the agent's displacement to a position $\mathcal{X}'$ can be accounted for by the rigid transformation $\mathcal{T}$ of entire space:

$$p' = \pi_{\mathcal{X}'}(x) = \pi_{\mathcal{X}}(\mathcal{T}(x)). \quad (5)$$

Assuming that proprioception unambigously defines the position of the exteroceptors in space

$$x = \pi_{\mathcal{X}}^{-1}(p)$$

and

$$s = \left(\sigma_{\mathcal{E}} \circ \pi_{\mathcal{X}}^{-1} \right)(p)$$

where the function $\left(\sigma_{\mathcal{E}} \circ \pi_{\mathcal{X}}^{-1} \right)$ is the sensorimotor contingency learned by the agent for every position $\mathcal{X}$ of itself and of the environment $\mathcal{E}$.

When the agent or the environment moves, a new sensorimotor contingency is established

$$s' = \left(\sigma_{\mathcal{E}'} \circ \pi_{\mathcal{X}'}^{-1} \right)(p') = \left(\sigma_{\mathcal{E}} \circ T \circ \mathcal{T}^{-1} \circ \pi_{\mathcal{X}}^{-1} \right)(p').$$

The agent learns the function φ linking the values p and p' such that $s = s'$, or

$$\left(\sigma_{\mathcal{E}} \circ \pi_{\mathcal{X}}^{-1} \right)(p) = \left(\sigma_{\mathcal{E}} \circ T \circ \mathcal{T}^{-1} \circ \pi_{\mathcal{X}}^{-1} \right)(p').$$

The function φ is not always defined uniquely since the mapping $\sigma_{\mathcal{E}}$ can be non-invertible. It can be inverted in the domain of its arguments if the environment is sufficiently rich, i.e., if the vector of exteroceptor outputs is different at every position of the exteroceptors within the range admitted by the proprioceptors. In this case

$$\varphi = \pi_{\mathcal{X}} \circ \mathcal{T} \circ T^{-1} \circ \pi_{\mathcal{X}}^{-1}. \quad (6)$$

It can be seen from the expression for the function φ that it simply gives a proprioceptive account of the relative rigid displacement $\mathcal{T} \circ T^{-1}$ of the environment and the agent. The functions φ are thus the agent's extensible rigid displacements, which are associated with the environment's rigid motion from $\mathcal{E}$ to $\mathcal{E}'$ and the agent's rigid motion from $\mathcal{X}$ to $\mathcal{X}'$. As is clear from equation 6, the function φ only depends on the transformations T and $\mathcal{T}$. In physical space, these transformations depend only on the displacements themselves and are independent of the initial positions of the agent and the environment, and of the content of the environment. Moreover, since the transformations T and $\mathcal{T}$ form Lie groups, the functions φ also inherit some group properties. For any two φ functions,

$$\varphi_1 = \pi_{\mathcal{X}} \circ \mathcal{T} \circ T_1^{-1} \circ \pi_{\mathcal{X}}^{-1} \quad \text{and} \quad \varphi_2 = \pi_{\mathcal{X}} \circ \mathcal{T}_2 \circ T_2^{-1} \circ \pi_{\mathcal{X}}^{-1}.$$

there exists a function φ_3 such that

$$\varphi_1 \circ \varphi_2 = \pi_{\mathcal{X}} \circ \mathcal{T}_1 \circ T_1^{-1} \circ \mathcal{T}_2 \circ T_2^{-1} \circ \pi_{\mathcal{X}}^{-1} = \pi_{\mathcal{X}} \circ \mathcal{T}_3 \circ T_3^{-1} \circ \pi_{\mathcal{X}}^{-1} = \varphi_3$$

where $\mathcal{T}_3$ and T_3 are transformations describing the total displacements of agent and of the environment, for which $\mathcal{T}_3 \circ T_3^{-1} = \mathcal{T}_1 \circ T_1^{-1} \circ \mathcal{T}_2 \circ T_2^{-1}$.

The functions φ do not form a group. This is because they are defined only on a subset of proprioceptive values, for which the exteroceptor outputs overlap before and after the shift. It may happen that the domain of definition of the function φ_3 is larger than that of $\varphi_1 \circ \varphi_2$ and hence the composition $\varphi_1 \circ \varphi_2$ is not one of the functions φ.

Up until this point we have assumed the existence of real physical space. Now, we would like to abandon this assumption, and only retain the conditions which allow the construction of the function φ. This gives us a list of requirements for the existence of "space-like" constructs. (1) There must be a variable x and functions $\sigma_{\mathcal{E}}$ and $\pi_{\mathcal{X}}$ such that the outputs of the extero- and proprioceptors can be described by the equations (2) and (4), and the function π must be invertible. The agent must be able to "act," i.e., induce changes in the variable x. (2) Moreover, there must exist (and be sufficiently often) changes of the environment $\mathcal{E} \to \mathcal{E}'$ and/or of the agent $\mathcal{X} \to \mathcal{X}'$ such that the equations (3) and (5) hold. The corresponding transformations T and/or $\mathcal{T}$ must be applicable to all environments $\mathcal{E}$ and external states of the agent $\mathcal{X}$, and they must form a group with respect to the composition operator.

Note that the requirement (2) does not presume that there are no other types of changes of the environment and/or of the agent. The agent will identify only the changes possessing such a property as *sensible rigid displacements* and will obtain the functions φ that correspond to them.

Also note again that here we do not assume the existence of space. We only make certain assumptions regarding the structure of the sensory inputs that the agent can receive.

The agent presented in **Figure 2** of the main text will only recognize translations as the spatial changes, because it can only translate its retina, and hence for this agent the variable x only includes the position of the retina in space, not its orientation.

One can imagine an agent that can stretch its retina in addition to translations and rotations. For such an agent, the variable x will include position, orientation, and stretching of the retina. If this agent can stretch its entire body, or if the environment has a tendency for such deformations, then stretching will be classified as a *sensible rigid displacement* similarly to translations and rotations.

One can also imagine an agent whose sensory inputs do not depend on physical spatial properties, but satisfy the requirements described above. Such an agent will develop a false notion of space, where it is not present. The description of such an agent is given below.

7. AUDIO AGENT

Here, we show that an agent can develop incorrect spatial knowledge, i.e., that does not correspond to physical space, if the conditions presented in the previous section are satisfied. The agent,
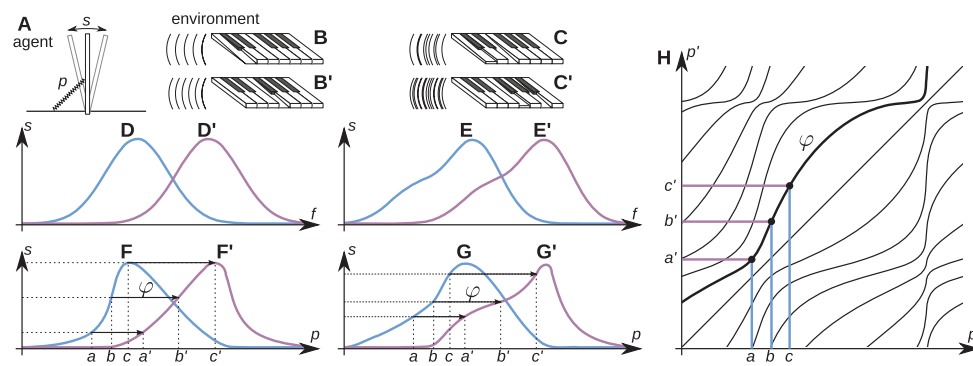
FIGURE 7 | A simple audio agent. The agent **(A)** consists of a rod attached to a support, oscillating in response to the acoustic waves **(B)**. The agent measures the intensity of this oscillation with an "exteroceptor" s. By exercising its muscle the agent can change the stiffness of the spring at the support and thus the eigenfrequency f of the rod, measured with the proprioceptor p. For an acoustic wave generated by a single note **(A)** the exteroceptor response s depends on the eigenfrequency f as shown in **(D)**. However, the agent measures the proprioceptive response p and not the eigenfrequency, and so only has access to the sensorimotor contingency **(F)**. The change of the note from **(B)** to **(B′)** results in a shift in the characteristics **(D)** to **(D′)** and in the establishment of a new sensorimotor contingency **(F′)**. The agent can learn the functions φ **(H)** by taking note of the coincidences in the exteroceptor output s. If a pair of notes is played **(C,C′)** with the dependencies **(E,E′)**, new sensorimotor contingencies are obtained by the agent **(G,G′)**, yet the same functions φ link them together. Reproduced with permission from Terekhov and O'Regan (2014) © 2016 IEEE.

inspired by Jean Nicod, inhabits the world of sounds (**Figure 7**). Its environment is a continuously lasting sine wave, or a chord (**Figure 7A**). The agent consists of a hair-cell, which oscillates in response to the acoustic waves (**Figure 7B**). The amplitude of this oscillation is measured by an exteroceptor s. The response is maximal if the frequency f of one of the sine waves coincides with the eigenfrequency of the hair-cell.

The agent can "scan" the environment by changing the stiffness at the cell's attachment point and thus its eigenfrequency, which is measured by the proprioceptor p. For the environment B, the dependency between the amplitude and the cell's eigenfrequency has the shape illustrated in **Figure 7D**. We assume that for any other note (like B′), the dependency between the amplitude and the cell's eigenfrequency remains the same, but shifted (**Figure 7D′**).

The agent does not know these facts. It only knows the dependency between exteroception s and proprioception p, which constitutes the sensorimotor contingency (**Figure 7F**) corresponding to the environment B. For a new note (B′), a new sensorimotor contingency F′ is established. Yet, as before, the agent notices that the outputs of the exteroceptor s coincide for certain values of p. It makes note of these coincidences and defines the functions φ corresponding to all changes of the notes (**Figure 7H**).

The same procedure applies if the agent faces a chord of two (C and C′) or more notes. Instead of changes in the pitch of a note, we now have transposition of the whole chord. The agent can discover that the same set of functions φ works for notes and for chords.

Although this agent is unable to move in space and although it only perceives continuous sound waves, it can nevertheless build the basic notions of space. However, these notions are "incorrect," in the sense that they do not correspond to actual physical space, but to the set of note pitches. The sensible rigid displacements for this agent correspond to transpositions of the chords. The unchanging medium is the musical scale, and the relative position of one chord with respect to an identical but transposed chord is

just the interval through which the chord has been transposed. For such an agent, a musical piece is somehow similar to what a silent film is for us: it is a sequence of objects (notes), appearing, moving around (changing pitch), and disappearing.

Using the formalism introduced above, we can say that for this agent, the spatial variable is frequency, f. For any given environment $\mathcal{E}$, which in this case is constituted by simultaneously played notes, the output of the exteroceptor s depends only on the eigenfrequency of the hair-cell, which can be measured using the same variable f. This means that the function $\sigma_{\mathcal{E}}(f)$ exists. The rigid shift of the environment $\mathcal{E}$ to $\mathcal{E}'$, which is the chord transposition, results just in frequency scaling: $\sigma_{\mathcal{E}'}(f) = \sigma_{\mathcal{E}}(kf)$. Evidently, these transformations form a group. Proprioception p signals the stiffness of the hair-cell, which is functionally related to its eigenfrequency, and hence the invertible function $\pi(f)$ also exists. As our auditory agent is unable to perform anything similar to rigid displacements, the function π does not depend on anything equivalent to the state $\mathcal{X}$ of our original simple agent (**Figure 1**).

The existence of the functions $\sigma_{\mathcal{E}}(f)$ and $\pi(f)$ fulfills the requirement (1) from the previous section. We can assume that music being played is just a piano exercise and hence the chords are often followed by their transposed versions. In this case there exist (and are sufficiently often) changes of the environment $\mathcal{E} \rightarrow \mathcal{E}'$, which correspond to a simple shift of all played notes by the same musical interval. These shifts evidently form a group, and hence the requirement (2) is also fulfilled. The fulfilment of these two requirements suffices for the existence of sensible rigid displacements and thus for basic spatial knowledge, described above.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

FUNDING

REFERENCES

Blakemore, C., and Cooper, G. F. (1970). Development of the brain depends on the visual environment. *Nature* 228, 477–478. doi:10.1038/228477a0

Bongard, J., Zykov, V., and Lipson, H. (2006). Resilient machines through continuous self-modeling. *Science* 314, 1118–1121. doi:10.1126/science.1133687

Ganguli, S., and Sompolinsky, H. (2012). Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annu. Rev. Neurosci.* 35, 485–508. doi:10.1146/annurev-neuro-062111-150410

Gloye, A., Wiesel, F., Tenchio, O., and Simon, M. (2005). Reinforcing the driving quality of soccer playing robots by anticipation. *Inf. Technol.* 47, 250–257. doi:10.1524/itit.2005.47.5_2005.250

Gordon, G., and Ahissar, E. (2011). "Reinforcement active learning hierarchical loops," in *Neural Networks (IJCNN), The 2011 International Joint Conference on (IEEE)*, (San Jose, CA: IEEE), 3008–3015.

Held, R., and Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *J. Comp. Physiol. Psychol.* 56, 872–876. doi:10.1037/h0040546

Hersch, M., Sauser, E. L., and Billard, A. (2008). Online learning of the body schema. *Int. J. HR* 5, 161–181. doi:10.1142/S0219843608001376

Hoffmann, M., Marques, H. G., Hernandez Arieta, A., Sumioka, H., Lungarella, M., and Pfeifer, R. (2010). Body schema in robotics: a review. *IEEE Trans. Auton. Ment. Dev.* 2, 304–324. doi:10.1109/TAMD.2010.2086454

Kaplan, F., and Oudeyer, P.-Y. (2004). "Maximizing learning progress: an internal reward system for development," in *Embodied Artificial Intelligence* (Springer), 259–270.

Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2005). "Empowerment: a universal agent-centric measure of control," in *Evolutionary Computation, 2005. The 2005 IEEE Congress*, Vol. 1 (Edinburgh: IEEE), 128–135.

Klyubin, E. S., Polani, D., and Nehaniv, C. L. (2004). "Tracking information flow through the environment: simple cases of stigmergy," in *Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems* (Boston: The MIT Press), 563–568.

Koos, S., Cully, A., and Mouret, J.-B. (2013). Fast damage recovery in robotics with the t-resilience algorithm. *Int. J. Rob. Res.* 32, 1700–1723. doi:10.1177/0278364913499192

Kuipers, B. (1978). Modeling spatial knowledge. *Cogn. Sci.* 2, 129–153. doi:10.1207/s15516709cog0202_3

Laflaquiere, A., Argentieri, S., Breysse, O., Genet, S., and Gas, B. (2012). "A non-linear approach to space dimension perception by a naive agent," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference* (Vilamoura: IEEE), 3253–3259.

MacKay, D. M. (1962). "Aspects of the theory of artificial intelligence," in *The Proceedings of the First International Symposium on Biosimulation Locarno, June 29 – July 5, 1960*, eds C. A. Muses and W. S. McCulloch (New York: Springer US), 83–104. doi:10.1007/978-1-4899-6584-4_5

Markram, H., Gerstner, W., and Sjöström, P. J. (2011). A history of spike-timing-dependent plasticity. *Front. Syn. Neurosci.* 3:4. doi:10.3389/fnsyn.2011.00004

Nicod, J. (1929). *The Foundations of Geometry and Induction*. London: Kegan Paul, Trench, Trubner & Co., Ltd.

O'Regan, J. K., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain. Sci.* 24, 939–972. doi:10.1017/S0140525X01000115

Philipona, D., O'Regan, J., and Nadal, J.-P. (2003). Is there something out there? inferring space from sensorimotor dependencies. *Neural Comput.* 15, 2029–2049. doi:10.1162/089976603322297278

Pierce, D., and Kuipers, B. J. (1997). Map learning with uninterpreted sensors and effectors. *Artif. Intell.* 92, 169–227. doi:10.1016/S0004-3702(96)00051-3

Poincaré, H. (1905). *Science and Hypothesis*. New York, NY: The Walter Scott Publishing Co., Ltd.

Roschin, V. Y., Frolov, A. A., Burnod, Y., and Maier, M. A. (2011). A neural network model for the acquisition of a spatial body scheme through sensorimotor interaction. *Neural Comput.* 23, 1821–1834. doi:10.1162/NECO_a_00138

Sigaud, O., Salaün, C., and Padois, V. (2011). On-line regression algorithms for learning mechanical models of robots: a survey. *Rob. Auton. Syst.* 59, 1115–1129. doi:10.1016/j.robot.2011.07.006

Terekhov, A. V., and O'Regan, J. K. (2014). "Learning abstract perceptual notions: the example of space," in *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2014 Joint IEEE International Conferences on*, p. 368–373.

von Uexküll, J. (1957). "A stroll through the worlds of animals and men: a picture book of invisible worlds," in *Instinctive Behavior: The Development of a Modern Concept*, ed. C. H. Schiller (New York: International Universities Press, Inc), 5–80.

Zhaoping, L. (2006). Theoretical understanding of the early visual processes by data compression and data selection. *Network* 17, 301–334. doi:10.1080/09548980600931995

Just Imagine! Learning to Emulate and Infer Actions with a Stochastic Generative Architecture

Fabian Schrodt* and Martin V. Butz

Cognitive Modeling, Department of Computer Science, University of Tübingen, Tübingen, Germany

Theories on embodied cognition emphasize that our mind develops by processing and inferring structures given the encountered bodily experiences. Here, we propose a distributed neural network architecture that learns a stochastic generative model from experiencing bodily actions. Our modular system learns from various manifolds of action perceptions in the form of (i) relative positional motion of the individual body parts, (ii) angular motion of joints, and (iii) relatively stable top-down action identities. By Hebbian learning, this information is spatially segmented in separate neural modules that provide embodied state codes and temporal predictions of the state progression inside and across the modules. The network is generative in space and time, thus being able to predict both, missing sensory information and next sensory information. We link the developing encodings to visuomotor and multimodal representations that appear to be involved in action observation. Our results show that the system learns to infer action types and motor codes from partial sensory information by emulating observed actions with the own developing body model. We further evaluate the generative capabilities by showing that the system is able to generate internal imaginations of the learned types of actions without sensory stimulation, including visual images of the actions. The model highlights the important roles of motor cognition and embodied simulation for bootstrapping action understanding capabilities. We conclude that stochastic generative models appear very suitable for both, generating goal-directed actions and predicting observed visuomotor trajectories and action goals.

Keywords: artificial neural networks, mental imagery, embodied simulation, sensorimotor learning, generative model, action understanding, action emulation, Bayesian inference

1. INTRODUCTION

It appears that humans are particularly good at learning by imitation, gaze following, social referencing, and gestural communication from very early on (Tomasello, 1999). Inherently, the observation of others is involved in all of these forms of social learning. Learning by imitation, for instance, is assumed to develop from pure mimicking of bodily movements toward the inference and emulation of the intended goals of others from about 1 year of age onward (Carpenter et al., 1998; Want and Harris, 2002; Elsner, 2007). Yet *how* are goals and intentions inferred from visual observations, and how does this facilitate the activation of the respective motor commands for imitation? The intercommunication between specific brain regions, which are often referred to as mirror neuron system or action observation network, has been suggested to enable this inference

of others' intentions and imitation of their behavior (Buccino et al., 2004; Rizzolatti and Craighero, 2004, 2005; Iacoboni, 2005, 2009; Iacoboni and Dapretto, 2006; Kilner et al., 2007). While a genetic predisposition may supply the foundation to develop such a system (Rizzolatti and Craighero, 2004; Ferrari et al., 2006; Lepage and Théoret, 2007; Bonini and Ferrari, 2011; Casile et al., 2011), its development – *per se* – seems to be strongly determined by social interaction (Meltzoff, 2007; Heyes, 2010; Nagai et al., 2011; Froese et al., 2012; Saby et al., 2012), sensorimotor experience, motor cognition, and embodiment (Gallese and Goldman, 1998; Catmur et al., 2007; Gallese, 2007a; Gallese et al., 2009). Due to observations such as the foregoing, cognitive science has recently undergone a pragmatic turn, focusing on the enactive roots of cognition (Engel et al., 2013).

Embodied cognitive states, according to Barsalous simulation hypothesis (Barsalou, 1999, 2008), are situated simulations that temporarily activate – or re-enact – particular events by means of a set of embodied modal codes. However, if mental states are grounded in own-bodily experiences and self-observations, how does the brain establish the correspondence to the observation of others in the first place? We have recently shown that this so-called correspondence problem [cf. Heyes (2001) and Dautenhahn and Nehaniv (2002)] can be solved by an embodied neural network model that is adapting to the individual perspectives of others (Schrodt et al., 2015). This model clustered sensorimotor contingencies and learned about their progress in a single competitive layer composed of cells with multimodal tuning, enabling it to infer proprioceptive equivalents to visual observations while taking an actors perspective.

In this paper, we propose a stochastic variant of the clustering algorithm, which we introduced in our previous work, that is generative in multiple, distributed domains. The system can be considered to develop several hidden Markov models from scratch and incorporates them by integrating conditional state transition probabilities statistically. It thereby learns an embodied action model that is able to simulate forward in time consistent visual-proprioceptive self-perceptions. This bodily grounded simulation is primed when observing biological motion patterns, leading to the ability to re-enact the observed behavior using the own embodied codes. Hence, our model supports the view that mental states are embodied simulations [cf. Gallese (2007b)] and provides an explanation to how the perception of others' actions can be consistently incorporated with the own action experiences when encoded at distributed neural sites.

Our model can be compared to an action observation network, in that it models the processing of (i) visual motion signals, believed to be processed in the superior temporal sulcus; (ii) spatiotemporal motor codes, which can be related to neural activities in the posterior parietal lobule and the premotor cortex, and (iii) compressed, intentional action codes, which have been associated with neural activities in the inferior frontal gyrus [see, e.g., Iacoboni (2005), Kilner (2011), and Turella et al. (2013)]. Accordingly, we train and evaluate a tripartite network structure, interpreting and referring to (i) relative positional body motion as *visual* biological motion stimuli, (ii) joint angular motion as *motor* codes, and (iii) action identities as *intentions* or *goals* in our experiments. In doing so, we focus on bodily movements,

including walking, running, and playing basketball, where the stimuli originate from motion captures of human subjects. Despite the simplicity of these stimuli, our results show that it is possible to identify compressed intention codes from observing biological motion patterns and to concurrently infer consistent motor emulations of observed actions using distributed, bodily grounded encodings. Analogously, actions can be simulated in visual and motor modalities when only an intention prior is provided, offering a possible explanation to how simulation processes may drive forth goal-directed and imitative behavior, and link it to social learning.

In the following, we refer to related work in Section 2 and specify the model architecture, including its modularized structure as well as the probabilistic learning and information processing mechanisms in Section 3. We then describe the motion capture stimuli, the bottom-up processing, and clarify the connection of the resulting perceptions to encodings involved in action understanding in Section 4. The model is evaluated on motion tracking data, showing action inference, completion, and imagination capabilities in Section 5. Finally, we discuss current challenges and future application options in Section 6.

2. RELATED WORK

Lallee and Dominey (2013) implemented a model that integrates low-level sensory data of an iCub robot, encoding multimodal contingencies in a single, 3D, and self-organizing competitive map. When driven by a single modal stimulus, this multimodal integration enables mental imagery of corresponding perceptions in other modalities. In accordance with findings from neuroscience, the modeled self-organizing map is topographic with respect to its discrete multimodal cell tunings. The states generated by our model can also be embedded in metric spaces. In contrast, however, our model encodes modal prototype vectors separately and activates them stochastically. This allows to encode multimodal perceptions without redundancies. Moreover, it enables the resolution of ambiguities over time by predictive interactions between the encoded modalities. Our results show that cells can be activated by multimodal perceptions without necessarily encoding multimodal stimuli locally, while moreover being able to encode specific actions by means of distributed temporal statistics.

Taylor et al. (2006) implemented a stochastic generative neural network model based on conditional restricted Boltzmann machines (RBMs). When trained on motion captures similar to those used in our evaluations, the model is able to reproduce walking and running movements as well as transitions between them in terms of sequences of angular postures. Although the encoding capacity of RBMs is theoretically superior in comparison to Markov state-based models because they encode multidimensional state variables, the experiments show the typical tradeoff of requiring considerably more training trials and randomized sampling. Our model is able to expand its encoding capacity on-demand and thus avoids both a sampling and frequency bias. Our model, nevertheless, accounts for scalability and encoding capacity since states are distributed over several Markov models. This enables to learn modal state transition densities locally and to

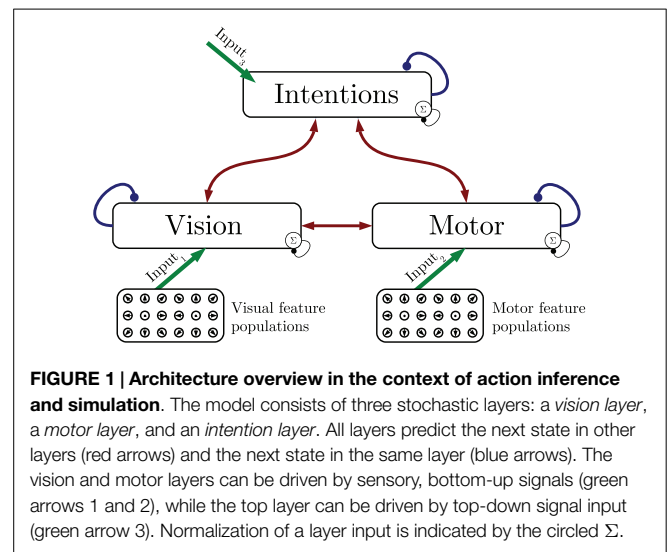reconcile them with sensory signals and cross-modal predictions as required.

Comparable to the realization by Baker et al. (2009) of a qualitative cognitive model suggested by Gergely et al. (1995), intention inferences in our model are based on Bayesian statistics given visually observed action sequences. In contrast, our model learns the sensorimotor contingencies that facilitate this inference without relying on specific behavioral rationality assumptions. Comparably, the intention priors in our model are statistically determined by assessing the own behavioral biases during an embodied training phase. Thereby, our experiments are based on the assumption that an observer expects an actor to behave in the same way they would behave – that is, by inferring cross-modal observation equivalences based on the own-bodily experiences – and thus essentially models the development of social cognition [cf. Meltzoff (2007)].

Similar to Friston et al. (2011), our neural network models action understanding by inferring higher level, compact action codes, given lower level sensory motion signals. However, in contrast to Friston et al. (2011), no motion primitives are provided, but they are learned in the form of intention clusters, which integrate sensory–motor information over space and time.

3. NEURAL NETWORK ARCHITECTURE

The stochastic generative neural model consists of several stochastic neural layers or modules, which process information in identical fashion. The layers can be arranged hierarchically and connected selectively. Each layer calculates a normalized, discrete probability density estimate for the determination of a state in a specific state space. Each neuron corresponds to a possible state, and the binary activation of a single cell corresponds to the determination of that state. The neurons are activated by developing and incorporating prototype tunings and temporal state predictions. Each neuron sends intramodular state transition predictions to the other neurons in the layer and cross-modular predictions to associated layers, such that the distributed states are able to develop self-preserving, generative temporal dynamics. The development of these predictions can be compared to predictive coding (Rao and Ballard, 1999) and results in a Hebbian learning rule similar to Oja's rule (Oja, 1989) as described in Section 3.3.

Figure 1 shows the particular network architecture developed here. Referring to the human action observation network, three layers of this kind interact with each other in a hierarchy of two levels: at the bottom level, a *vision layer* processes bottom-up visual motion cues and predicts the continuation of this visual motion over time as well as corresponding action intentions and motor codes. Further, a *motor layer* processes bottom-up proprioceptions of joint angular motion and predicts the continuation of these signals over time as well as corresponding action intentions and visual motion. Finally, at the top level, an *intention layer* encodes the individual actions for which the system is trained on, predicts possible action transitions over time, and top-down the corresponding vision and motor layer states that may be active during a particular action. Hence, at the bottom level, top-down and generative activities are fused
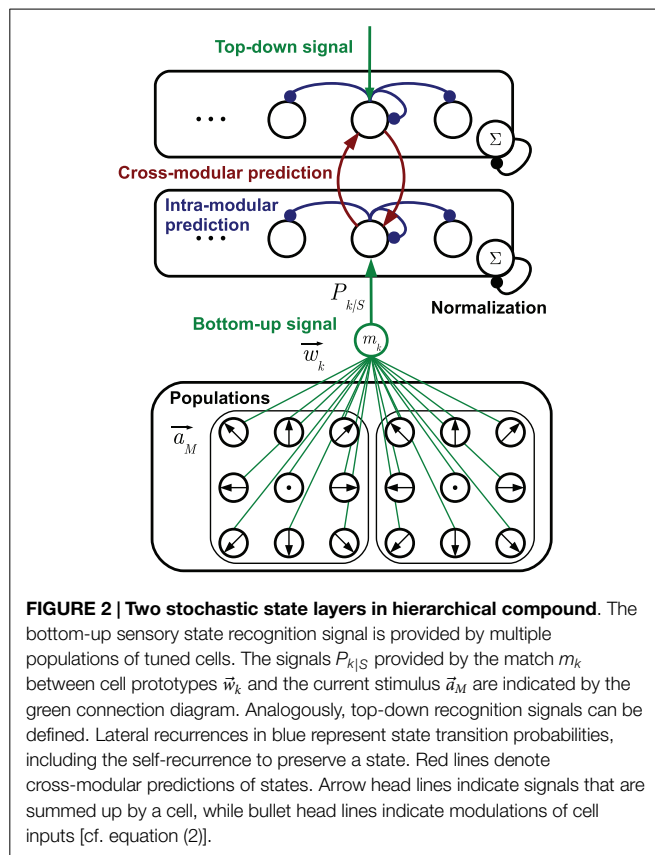

FIGURE 1 | Architecture overview in the context of action inference and simulation. The model consists of three stochastic layers: a *vision layer*, a *motor layer*, and an *intention layer*. All layers predict the next state in other layers (red arrows) and the next state in the same layer (blue arrows). The vision and motor layers can be driven by sensory, bottom-up signals (green arrows 1 and 2), while the top layer can be driven by top-down signal input (green arrow 3). Normalization of a layer input is indicated by the circled Σ.

with bottom-up sensory signals, in common with the intramodular and cross-modular predictions generated by the bottom layers themselves. In a context where each bottom module represents a specific modality, the intramodular predictions can be considered to represent the expected state progression in the respective modality, while cross-modular predictions implement cross-modal inferences. The cross-modular predictions enable the inference of motor and intention codes from visual observations during action observation, where only visual motion cues are available.

The streams of sensory information are assumed to be provided by populations of locally receptive cells with tuning to specific stimuli, which is in accordance with findings in neuroscience (Pouget et al., 2000). These populations essentially forward the information by means of a full connection to the bottom stochastic layer that reflects the corresponding modality. Section 4 elaborates further on how the respective perceptions and stimuli are encoded and how they can be related to an action observation network. This encoding has been published recently as part of a perspective-inference model given dynamic motion patterns (Schrodt et al., 2015). The following sections thus focus on the stochastic neural layers on top of the populations.

3.1. Stochastic Neural Layers

Each stochastic neural layer learns a discrete, prototypic representation of the provided sensory input information. To do so, the layer grows a set of cells on demand with distinct sensory tunings. The recruitment of cells and adaptation of prototypes is accomplished by unsupervised mechanisms as explained in Section 3.2. Each cell in a layer learns predictions of the temporal progress of these prototypic state estimates in the layer. Furthermore, each cell learns to predict the cell activations that may be observed in other, associated layers, which is explained in Section 3.3. An exemplary stochastic neural layer connected to another layer in this way, together with the neural populations that forward sensory signals is shown in **Figure 2**. In the following,

FIGURE 2 | Two stochastic state layers in hierarchical compound. The bottom-up sensory state recognition signal is provided by multiple populations of tuned cells. The signals $P_{k|S}$ provided by the match m_k between cell prototypes $\vec{w}_k$ and the current stimulus $\vec{a}_M$ are indicated by the green connection diagram. Analogously, top-down recognition signals can be defined. Lateral recurrences in blue represent state transition probabilities, including the self-recurrence to preserve a state. Red lines denote cross-modular predictions of states. Arrow head lines indicate signals that are summed up by a cell, while bullet head lines indicate modulations of cell inputs [cf. equation (2)].

the determination of states and incorporation of predictions is formalized.

The layers in our model simplify competitive neural processes such that only a single cell in each layer is activated at the same time. Cell activations are binary and represent the event that a specific state in the corresponding state space is determined. This is comparable to a winner-takes-all approach [cf. Grossberg (1973), for evaluations]. However, the determination of the state in each layer depends on a fusion of predictive intramodular and cross-modular probabilities and sensory state recognition probabilities. By stochastic sampling, a single cell is selected as *competition winner* in each time step, where the winning probability is determined by the fused inputs to each cell. In the process, the input vector to a layer depicts a discrete probability density for the stochastic event of observing a particular state. For this reason, each layer uses a specific normalization of incoming signals that ensures that the all signals sum up to 1.

We denote cells inside a layer by an index set M and cells outside by an index set N. The binary output $x_k(t)$ of a state cell indexed $k \in M$ is determined by the normalized probability term

$$X_k(t) = P(x(t) > x_j(t) \ \forall j \neq k, j \in M)$$
$$= \frac{\text{net}_k(t)}{\sum_{j \in M} \text{net}_j(t)} \tag{1}$$

where $X_k(t)$ denotes the winning event probability, and $x_k(t) \in \{0,1\}$ denotes the realization of this probability or abstract, binary cell activation calculated by stochastic sampling

at time step t. The input $\text{net}_k(t)$ to the cell k is provided by the probability fusion

$$\text{net}_k(t) = \big(P_{k|S}(t) + P_{k|C}(t)\big) \cdot P_{k|I}(t) \tag{2}$$

where $P_{k|S}(t)$ is a *sensory* (S) recognition signal depicting the probability that the state k is considered the current observation given sensory inputs, $P_{k|I}(t)$ is the *intramodular* (I) prediction of the successor state, and $P_{k|C}(t)$ is the *cross-modular* (C) prediction of the succession, defined by

$$P_{k|I}(t) = \prod_{i \in M} 1 - x_i(t-1) \cdot (1 - P(x_k(t) = 1 | x_i(t-1) = 1))$$
$$\tag{3}$$

$$P_{k|C}(t) = \sum_{j \in N} x_j(t-1) \cdot P(x_k(t) = 1 | x_j(t-1) = 1) \tag{4}$$

Taken together, equation (2) firstly fuses probabilistic sensory recognition signals with probabilistic cross-modular predictions coming in from the last winner cells of other layers. Then, it restricts the activation of cells to probabilistic intramodular predictions propagated from the last winner cell in the layer to all potential successors (including the last winner itself), as indicated in **Figure 2**.

The sensory recognition probability $P_{k|S}(t)$ is also responsible for clustering the sensory streams into discrete, prototypic states. In the following, we explain the segmentation by unsupervised Hebbian learning.

3.2. Segmentation and Recognition of Population-Encoded Activations

For generating the above binary stochastic cells, we use an instar algorithm that is capable of unsupervised segmentation of normalized vector spaces similar to Grossberg's Adaptive Resonance Theory (Grossberg, 1976a,b,c). In contrast, our approach provides state recognition probabilities and can thus be applied to implement non-deterministic learning and recognition. Another difference to common implementations is that cell prototypes are created on demand and initialized with zero vectors.

We define the sensory recognition probability $P_{k|S}(t)$ of a state $k \in M$ as a function of the congruence or match $m_k(t)$ between a state cell's prototype vector $\vec{w}_k$ and the current activation vector $\vec{a}_M(t)$ jointly provided by all population cells. The concatenated population activation dedicated to a state layer is assumed to be normalized to length 1. Since the model is designed for a separate learning and testing phase, we provide separate recognition functions, assuming full sensory confidence during training, and some sensory uncertainty during testing, which generally means observing previously unseen data. During training, this assumption inevitably results in the sensory recognition of the best matching state via

$$P_{k|S}^{\text{training}}(t) = \begin{cases} 1 & \text{if } m_k(t) \geq m_l(t) \ \forall l \in M \\ 0 & \text{else} \end{cases} \tag{5}$$

as well as a sensory recognition that is distributed over all states during testing, which we define by

$$P_{k|S}^{\text{testing}}(t) = \beta \cdot \frac{2}{1 + \exp(-\kappa(m_k(t) - 1))} \tag{6}$$

where κ denotes an uncertainty measure for sensory data, and β denotes the maximum sensor confidence. The prototype match to the current stimulus is described by

$$m_k(t) = \begin{cases} \frac{\vec{a}_M(t) \odot \vec{w}_k}{\|\vec{w}_k\|} & \text{if } k \text{ is } recruited \\ \theta & \text{if } k \text{ is } free \end{cases} \in [-1, 1] \quad (7)$$

where $\odot$ denotes the scalar product, such that the match function is based on the angular match between the normalized prototype vector $\vec{w}_k$ encoded in cell k and the current normalized stimulus $\vec{a}_M(t)$. Each layer expands its capacity on demand, comparable to Growing Neural Gas by Fritzke (1995). When a cell has fired a sensory recognition signal $[P_{k|S}(t) = 1]$ once during training, it is converted from a *free cell* to a *recruited cell* in the sense that its prototype vector is adapted from zero to the current stimulus [following the learning rule in equation (8)]. The match of a free cell is fixed to θ, such that when no cell match is greater than θ, the free pattern is recruited and another free cell is created with zero vector prototype. Thus, we call θ the *recruitment threshold* in the following. Assuming a small learning rate, we can ensure that each training input is encoded in the network with a tolerance mismatch of θ, irrespective of the amount of data, the presentation order, or frequency. Further, it was suggested previously that adding noise to the match function introduces a specific degree of noise robustness to this segmentation algorithm during training (Schrodt et al., 2015).

Prototype vectors of cells are trained to represent the current population activation using the Hebbian inspired instar learning rule:

$$\nabla \vec{w}_k(t) = \eta_s \cdot x_k(t) \cdot (\vec{a}_M(t) - \vec{w}_k(t)) \quad (8)$$

where η_s denotes the spatial learning rate. Since learning is gated by the binary cell realization $x_k(t)$, only the prototype of the winner cell is adapted.

During testing, the sensory recognition function [equation (6)] ensures the distribution of sensory state recognition probabilities over all stochastic cells rather than a single one to account for sensory uncertainty. Perfectly matching cells are recognized with probability β (before normalization), whereas the probability to recognize states not perfectly in the center of the stimulus decreases in dependency on κ and the mismatch. This means also that when no learned prototype matches sufficiently well during testing, the sensory recognition distribution becomes nearly uniform, such that intramodular and cross-modular predictions gain a relatively strong influence on the determination of the current state [cf. equation (2)]. Therefore, the network is able to dynamically switch from a bottom-up driven state recognition to a forward simulation of the state progression when sensory information is unknown or uncertain. In the following, we detail how intramodular and cross-modular predictions can be learned by a Hebbian learning rule that is equivalent to Bayesian inference.

3.3. Learning Intramodular and Cross-Modular Predictions

Upon winning, a cell learns to predict which observations will be made next in the same and in other layers. This is realized by asymmetric bidirectional recurrences between cells in a layer, representing the intramodular predictions $P_{k|I}(t)$, and between cells of two layers, representing the cross-modular predictions $P_{k|C}(t)$. Intramodular recurrences propagate the state transition probability from the last winner to all cells in the same layer and thus implement a discrete-time Markov chain, where Markov states are learned from scratch during the training procedure. Cross-modular connections bias the state transition probability density in other layers, given the current sensory observation, by means of temporal Bayesian inference.

Taken together, in a fully connected architecture, intramodular and cross-modular state predictions are represented by a full connection between all state cells in the network (including self-recurrences). These connections generally encode conditional probabilities for the subsequent observation of specific states. They can be learned by Bayesian statistics, which would result in asymmetric weights.

$$w_{ij}(t) = P(x_j(t) = 1 | x_i(t-1) = 1) = \frac{\sum_t x_i(t-1) \cdot x_j(t)}{\sum_t x_i(t-1)} \quad (9)$$

$$w_{ji}(t) = P(x_i(t) = 1 | x_j(t-1) = 1) = \frac{\sum_t x_i(t) \cdot x_j(t-1)}{\sum_t x_j(t-1)} \quad (10)$$

To derive a neurally more plausible learning rule to train a weight from cell i to cell j, we transpose the derivative of this formula with respect to time:

$$\frac{\partial w_{ij}(t)}{\partial t}$$

$$= \frac{\frac{\partial \sum_t x_i(t-1) \cdot x_j(t)}{\partial t} \sum_t x_i(t-1) - \frac{\partial \sum_t x_i(t-1)}{\partial t} \sum_t x_i(t-1) \cdot x_j(t)}{\left(\sum_t x_i(t-1)\right)^2}$$

$$= \frac{x_i(t-1) \cdot x_j(t) \cdot \sum_t x_i(t-1) - x_i(t-1) \cdot \sum_t x_i(t-1) \cdot x_j(t)}{\left(\sum_t x_i(t-1)\right)^2}$$

$$= \frac{x_i(t-1) \cdot x_j(t) - x_i(t-1) \cdot w_{ij}(t)}{\sum_t x_i(t-1)}$$

$$= \frac{x_i(t-1) \left(x_j(t) - w_{ij}(t)\right)}{\sum_t x_i(t-1)}$$

$$= \eta_p \cdot x_i(t-1) \cdot (x_j(t) - w_{ij}(t)), \eta_p = \frac{1}{\sum_t x_i(t-1)} \quad (11)$$

With the predictive learning rate η_p set constant, this is a temporal variant of Oja's associative learning rule (Oja, 1989), also referred to as outstar learning rule. Thus, this form of Hebbian learning is equivalent to Bayesian inference under the assumption of a learning rate that decays inversely proportional to the number of activations of the preceding cell i. In this case, each cell calculates the average of all observed (temporally) conditional probability densities in the same and other layers. However, since the states are adapted simultaneously with the learning of state conditionals, it is advantageous to implement a form of forgetting. Hence, we define the learning rate by $\eta_p = \frac{1}{(\sum_t x_i(t-1))^\alpha}$, where $\alpha < 1$ implements forgetting. All state predicting weights w_{ij} are initialized equally to represent multiple uniform distributions, and adapt in accordance with learning rule 11.

The capability of simulating distributed state progressions, also without sensory stimulation, follows from the stochastic selection of cell activations based on the learned, conditional state predictions. As a result of the bidirectional connections, the model becomes able to infer momentarily or permanently unobservable states and to mutually synchronize, or keep consistent, activations in the respective layers. By pre-activating a subset of cells in a layer, also a subset of learned state sequences can unfold. In the context of actions, this leads to the ability to synchronously simulate the state progression that corresponds to one of multiple encoded bodily movements in the vision and motor layers when biased top-down by a constant intention signal. The probability fusion in equation (2) accounts for an approximation of the respective, multi-conditional state probabilities. In the following section, we describe in further detail the application of this model to action understanding and the respective stimuli used in our evaluations.
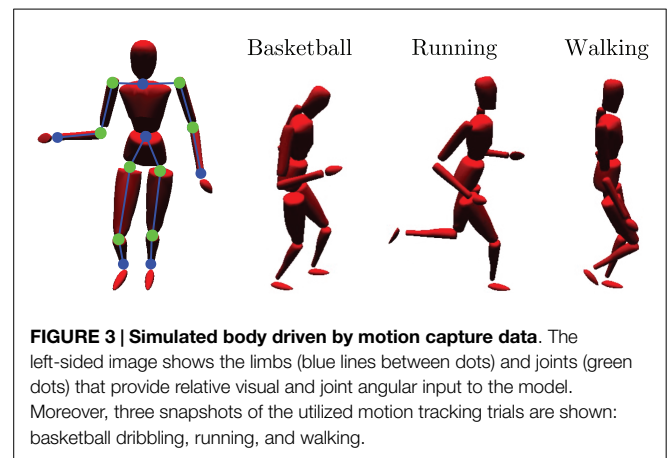
4. MODELING ACTION OBSERVATION

The focus of this paper lies on the learning of an embodied, distributed, and multimodal model of action understanding, which involves bottom-up as well as top-down and generative processes. It consists of three stochastic layers, each modeling codes and processes that are believed to be involved in action observation, the inference of goals, and respective motor commands that facilitate the emulation of observed actions. The first layer comprises *visual biological motion* patterns. The second layer encodes the corresponding joint angular *motor perceptions*. Accordingly, the model includes two groups of modal input populations, which encode visual and proprioceptive stimuli. Moreover, we include an amodal or multimodal intrinsic representation of *action intentions*. These codes are believed to be represented at distributed neural sites. It is typically assumed that action goals and intentions are encoded inferior frontally, motor codes and plans posterior parietally, and biological, mainly visually driven motion patterns in the superior temporal sulcus [cf. Iacoboni (2005), Kilner (2011), and Turella et al. (2013)]. Inferences and synchronization processes between these neural sites are modeled by cross-modular state predictions between the layers in the network, while the intramodular predictions restrict the state progression to the experienced, own-bodily contingencies. **Figure 1** shows an overview of the implemented learning architecture in this context.

In the following, we describe the bottom-up processing chain of our model referring to psychological and neuroscientific evidence. We start with the simulation environment and the motion capture data format that provides the respective stimuli for our evaluations. Subsequently, we focus on important key aspects for the recognition of biological motion, their implications, and implementation in the model. Finally, we describe how the resulting perceptions are interpreted in the context of different modalities involved in action perception, inference, and emulation.

4.1. Motion Captures and Data Representation

We evaluate our model making use of the CMU Graphics Lab Motion Capture Database (http://mocap.cs.cmu.edu/). Recordings from subjects performing three different cyclic movements

FIGURE 3 | Simulated body driven by motion capture data. The left-sided image shows the limbs (blue lines between dots) and joints (green dots) that provide relative visual and joint angular input to the model. Moreover, three snapshots of the utilized motion tracking trials are shown: basketball dribbling, running, and walking.

(*walking*, *running*, and *basketball dribbling*) in three trials each were utilized, as shown in **Figure 3**. For each movement, we chose a short, cyclic segment of the first trial as the training set and the other two, full trials as the testing set. In this way, the training set was rather idealized, while the testing set contained more information which, although inside the same action classes, strongly differed to the training data. The motion tracking data were recorded with 12 high-resolution infra-red cameras at 120 Hz using 41 tracking markers attached to the subjects. The resulting 3D positions were then matched to separate skeleton templates for learning and testing to obtain series of *joint angular postures* and coherent *relative joint positions*.

In the experiments, we chose the time series of 12 of the calculated relative joint positions as input to the visual processing pathway of the model. We selected the start and end points of the left and right upper arm, forearm, upper and lower leg, shoulder, and hip joints relative to the waist, as shown in **Figure 3**. Each was encoded by a three-dimensional Cartesian coordinate. As input to the motor pathway, we chose the calculated joint angles of 8 joints, each encoded by a one- to three-dimensional radian vector, depending on the degrees of freedom of the respective joint. We selected the left and right hip joints, knee joint, shoulder joints, and the elbow joints, resulting in 16 DOF overall. A map of the inputs at a single, exemplary time step is shown in **Figure 3**. The visual and motor pathways are neural substructures of the here proposed model and preprocess the raw data as described in the following.

4.2. Aspects of Biological Motion and Preprocessing

Giese and Poggio (2003) summarize critical properties of the recognition of biological motion from visual observations, such as selectivity for temporal order, generality, robustness, and view dependence. First, scrambling the temporal order in which biological motion patterns are displayed typically impairs the recognition of the respective action. This temporal selectivity is realized in our model by learning temporally directed state predictions. Second, biological motion recognition is highly robust against spatiotemporal variances (such as position, scale, and speed), body morphology and exact posture control, incomplete

representations (such as point-light displays), or variances in illumination. We model these generalization capabilities by means of (i) the usage of simplified forms of representation of biological motion stimuli as described above, (ii) the extraction of invariant and valuable information in a neural preprocessing stage, and (iii) the simulation of observed motion with the own embodied encodings. Third, the recognition performance decreases with the amount of rotation an action is perceived from with respect to common perspectives. The prototypic cells in our network also respond to specific, learned views of observed movements. However, the preprocessing of our model is able to also infer and adapt to observed perspectives to a certain degree.

This neurally deployed preprocessing is a part of the model that is not detailed in this paper. To summarize, the extraction of relevant information results in fundamental spatiotemporal invariances of the visual perception to scale, translation, movement speed, and body morphology. This is achieved by (i) exponential smoothing to account for noise in the data, (ii) calculation of the velocity, and (iii) normalization of the data to obtain the relative motion direction of each relative feature processed [see Schrodt and Butz (2014) and Schrodt et al. (2014a,b, 2015) for details]. For reasons of consistency, both the visual and motor perceptions are preprocessed in this manner. As to visual perception, the preprocessing stage is able to account also for invariance to orientation by means of active inference of the perspective an observed biological motion is perceived from. Compensating for the perspective upon observation solves the correspondence problem, which can be considered a premise for the ability to infer intrinsic action representations of others using the own, embodied encodings, as detailed in our previous work. As a matter of focus, however, we neglect the influence of orientation in the following experiments, meaning that the orientation of the learned and observed motions was identical.

Visual stimuli preprocessed in this manner are represented by a number of neural populations, each encoding the spatially relative motion direction of a specific bodily feature. Consequently, each cell in a population is tuned to a specific motion direction of a limb. Following this, the visual state layer accomplishes a segmentation of the concatenation of all visual population activations into whole-body, directional motion patterns. Analogously, the directions of changes in the joint angles are represented by populations and segmented into whole-body motor codes. In the following, we draw a comparison of this visuomotor perspective and our representation of intention codes to findings in neuroscience and psychology.

4.3. Visuomotor Perspective and Intentions

The superior temporal sulcus is particularly well known for encoding (also whole-body) biological motion patterns (Bruce et al., 1981; Perrett et al., 1985; Oram and Perrett, 1994) and has been considered to provide important visual input for the development of attributes linked with the mirror neuron system (Grossman et al., 2000; Gallese, 2001; Puce and Perrett, 2003; Ulloa and Pineda, 2007; Pavlova, 2012; Cook et al., 2014). Visual motion cues are necessary and most critical for the recognition of actions (Garcia and Grossman, 2008; Thurman and Grossman, 2008). As initially shown by Johansson (1973), the perception

of point-like bodily landmarks in relative motion is sufficient in this process. Thus, we assume that the above relative directional motion information can be perceived visually and is sufficient for action recognition. In contrast, joint angular motion cannot be perceived directly from such minimal visual information, which particularly applies to inner rotations of limbs. Thus, we assume that the directional angular limb motion is perceived proprioceptively. In the context of actions, we consider a prototype of such whole-body joint angular motion a *motor code*. Similar motor codes are assumed to be activated during the observation of learned movements (Calvo-Merino et al., 2005) and may be found in posterior parietal areas and related premotor areas (Iacoboni, 2005; Friston et al., 2011; Turella et al., 2013).

Further, in the context of the mirror neuron system, intentional structures can be assumed to be encoded in the inferior frontal gyrus (Iacoboni, 2005; Kilner, 2011; Turella et al., 2013). We simplify these intention codes by top-down, symbolic representations of specific actions. For the following experiments, we define three binary intentions in line with the motion tracking recordings explained before (basketball, running, and walking). Due to this symbol-like nature, the resulting intention layer cells can also be considered action classes or labels, while the derivation of intentions can be considered an online classification of observed bodily motion given visual cues. Since intentions are provided during training, the intention state cells and their predictions can be considered to develop by supervised training of action labels. However, all state variables are segmented using the unsupervised algorithm as described in Section 3.2.

During the observation of others, neither information about their proprioceptions nor their intentions are directly accessible. According to the embodied simulation hypothesis, the developing embodied states can nevertheless be inferred when observing others (Barsalou, 1999, 2008; Calvo-Merino et al., 2005). Hence, in the following experiments, we evaluate the inference and embodied simulation capabilities of our model.

5. EVALUATIONS

In the following experiments, we evaluate (a) the embodied learning of modal prototypes and predictions by means of the segmentation of different streams of information into prototypic state cells, (b) the resulting ability to infer intentions and motor states upon the observation of others' actions, and (c) the model's capability to simulate movements without sensory stimulation, keeping visual and motor states consistently. For all of the experiments, we chose the parameterization $\eta_s = 0.01$, $\alpha = 0.9$, $\beta = 0.5$, $\kappa = 16$, and $\theta = 0.85$ unless stated otherwise.

5.1. Experiment 1: Learning a Sensorimotor Model Mediated by Intentions

In the first experiment, we show how state cells develop from scratch given streams of relative visual and motor motion input. As shown in **Figure 4**, all layers are driven by data, assuming maximum sensory confidence and thus disabling the influence of predictions. Training consisted of learning perfectly cyclic motion tracking snippets: first, a 115 time steps or 0.96-s basketball trial where a single dribble and 2 footsteps were performed was shown

11 times in succession, resulting in 1265 time steps of training. Then, a 91 time steps or 0.75-s running trial performing 2 foot-steps was shown 14 times, resulting in 1274 frames. Finally, a 260 time steps or 2.17-s walking trial performing 2 steps was shown

5 times repeatedly, resulting in 1300 frames. The training data thus consisted of 3.88 s of unique data samples. The whole cyclic repetition of these trials was streamed into the model five times, while recruiting states, learning state prototypes and the resulting intra- and cross-modular predictions.

Figure 5 shows the recruitment of five visual and three motor state cells from scratch and the respective match to the driving stimuli in the example of a recruitment threshold $\theta = 0.1$. Because of the cyclic nature of the trained movements, the activations of those states form cyclic time series. The recruitment threshold θ basically defines the discretization of the state spaces. Hence, the higher the recruitment threshold θ, the more states develop, as concluded in **Table 1**. Note that learning was deterministic in these settings, which means that (a) adapted weights were not initialized randomly, but with a zero vector and (b) we assumed full sensor confidence such that the probability to recognize a state is a binary function. In consequence, there was no variance in the developing states.

Figure 5 also indicates that non-disjunct state encodings develop for the three different movements: only one of the states is recognized exclusively during the perception of a specific movement. Thus, classifications of movements are barely possible using
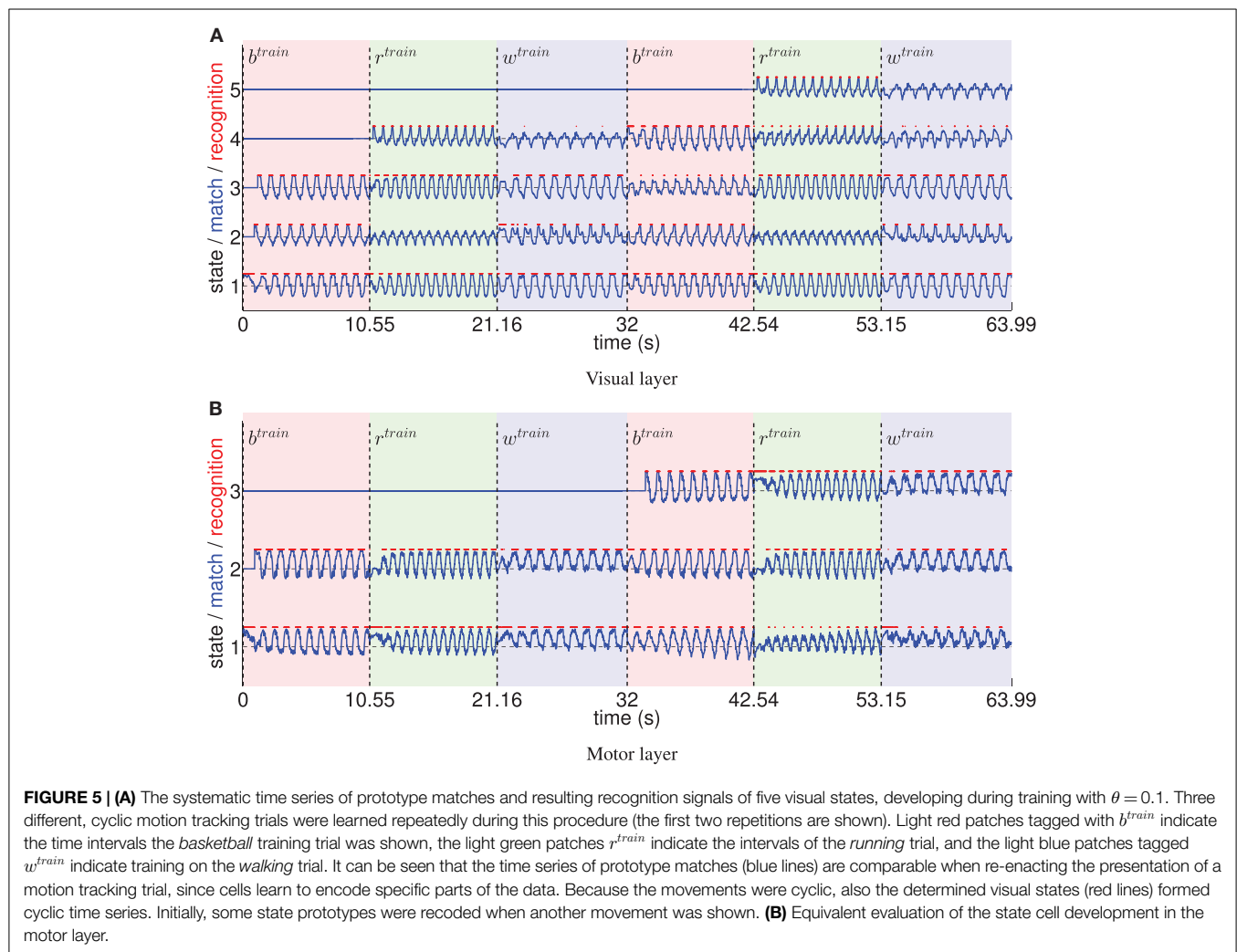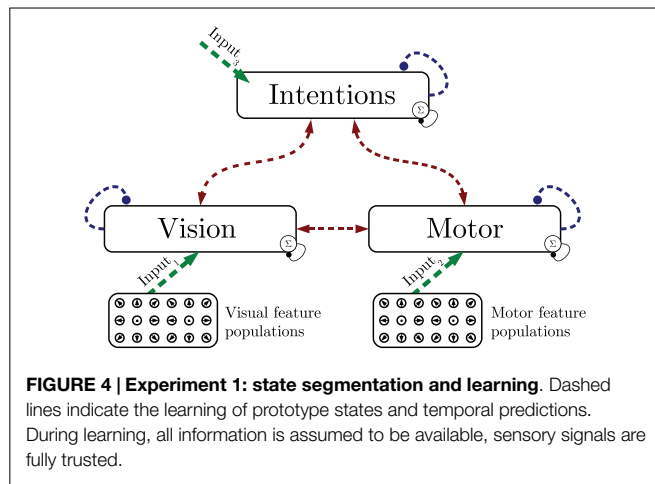

FIGURE 4 | Experiment 1: state segmentation and learning. Dashed lines indicate the learning of prototype states and temporal predictions. During learning, all information is assumed to be available, sensory signals are fully trusted.

FIGURE 5 | (A) The systematic time series of prototype matches and resulting recognition signals of five visual states, developing during training with $\theta = 0.1$. Three different, cyclic motion tracking trials were learned repeatedly during this procedure (the first two repetitions are shown). Light red patches tagged with b^{train} indicate the time intervals the *basketball* training trial was shown, the light green patches r^{train} indicate the intervals of the *running* trial, and the light blue patches tagged w^{train} indicate training on the *walking* trial. It can be seen that the time series of prototype matches (blue lines) are comparable when re-enacting the presentation of a motion tracking trial, since cells learn to encode specific parts of the data. Because the movements were cyclic, also the determined visual states (red lines) formed cyclic time series. Initially, some state prototypes were recoded when another movement was shown. **(B)** Equivalent evaluation of the state cell development in the motor layer.

TABLE 1 | Overview of the number of developing states during learning in dependency on θ and the resulting classification performances during observation of movements not seen during training.

θ	Layer	No. of developing states	Correct classifications (%)			Classifier confidence (%)		
			Basketball	Running	Walking	Basketball	Running	Walking
0.1	Visual	6	21.49	92.36	96.14	44.71	44.65	47.00
	Motor	3						
	Intentions	3						
0.3	Visual	8	60.31	98.28	95.43	50.90	50.95	50.75
	Motor	7						
	Intentions	3						
0.5	Visual	16	51.64	99.66	99.71	53.97	59.88	65.99
	Motor	15						
	Intentions	3						
0.7	Visual	31	43.26	98.51	99.24	60.83	67.97	75.79
	Motor	37						
	Intentions	3						
0.85	Visual	72	53.02	99.02	99.18	65.10	73.50	80.73
	Motor	107						
	Intentions	3						

Correct classification denotes the percentage of time steps the maximally likely intention output corresponded to the actually shown movement. The classifier confidence shows the average inferred probability of the maximally likely intention during testing.

Bayesian statistics with such a low recruitment threshold. Hence, in the following section, we examine the influence of increasing the visual and motor state granularity on the model's ability to infer movement classes.

5.2. Experiment 2a: Inference of Intentions upon Observation

For the classification of movements, or in this context, for the inference of intentions, the distinctness of the state structures with respect to the movements developing during training plays a major role. Since the information the state cells are encoding in their prototype vector is hard to visualize, we calculated the average pixel snapshot of the simulation display for each state while it was recognized [using an averaging formula analogously to equation (11)]. Basketball movements were displayed in red, running movements in green, and walking movements in blue. Consequently, if only a single state was created to represent all of the training data, the resulting state snapshot would show a mixture of all postures included in all of the movements, while overlapping postures would be black and non-overlapping postures would be colored. On the contrary, a state cell that was recognized only at a single time step during training would result in a snapshot showing only the corresponding posture in the respective color of the movement. Hence, the color of the snapshots can be considered a qualitative measure for the distinctness of states with respect to the three movements. Also, each snapshot shows the segments of the movements a state cell responds to and thus the model's "imagination" of the movement when modalities are inferred or simulated. **Figure 6** shows exemplar snapshots of cells created during the training phases using different recruitment thresholds θ. As expected, higher thresholds lead to the creation of movement-exclusive states.

To evaluate the influence of the multimodal state segmentation on the model's ability to infer intentions and to test for

FIGURE 6 | Average simulation display while specific sensory states were observed. (A) Displays an example snapshot of a state created with learning threshold $\theta = 0.3$. It shows that the state was recognized non-exclusively, that is, for each of the learned movements. The disjunction of patterns improves when increasing the threshold, such that for $\theta = 0.85$, movements and specific parts of the movements are clearly identifiable by observing specific states [snapshots **(B–D)**].

generalization at the same time, we measured the influence of θ on the correctness of the inferred values and the model's confidence, when different movements were presented after training. As indicated above, the testing set did not contain the motion tracking trials trained on. Rather, it contained two other basketball trials of 4.39 and 3.2 s, two other running trials of 3.56 s each, and two other running trials of 1.15 and 1.27 s. The testing data thus consisted of 17.13 s of unique data samples. Some trials included motion segments very different from the learned movements. Particularly, the basketball testing trials contained segments where the subject stood still and was lifting the ball or segments where the dribbling was incongruent with the footstep cycle, whereas the model was only trained on a single, congruent basketball dribbling snippet. Also, as indicated in **Figure 7**, only the visual modality was fed into the network during testing trials, which accounts for the fact that intentions and also motor commands are not directly observable during observation of actions. Note that the model did not obtain information about the time step when a new movement was shown during testing.
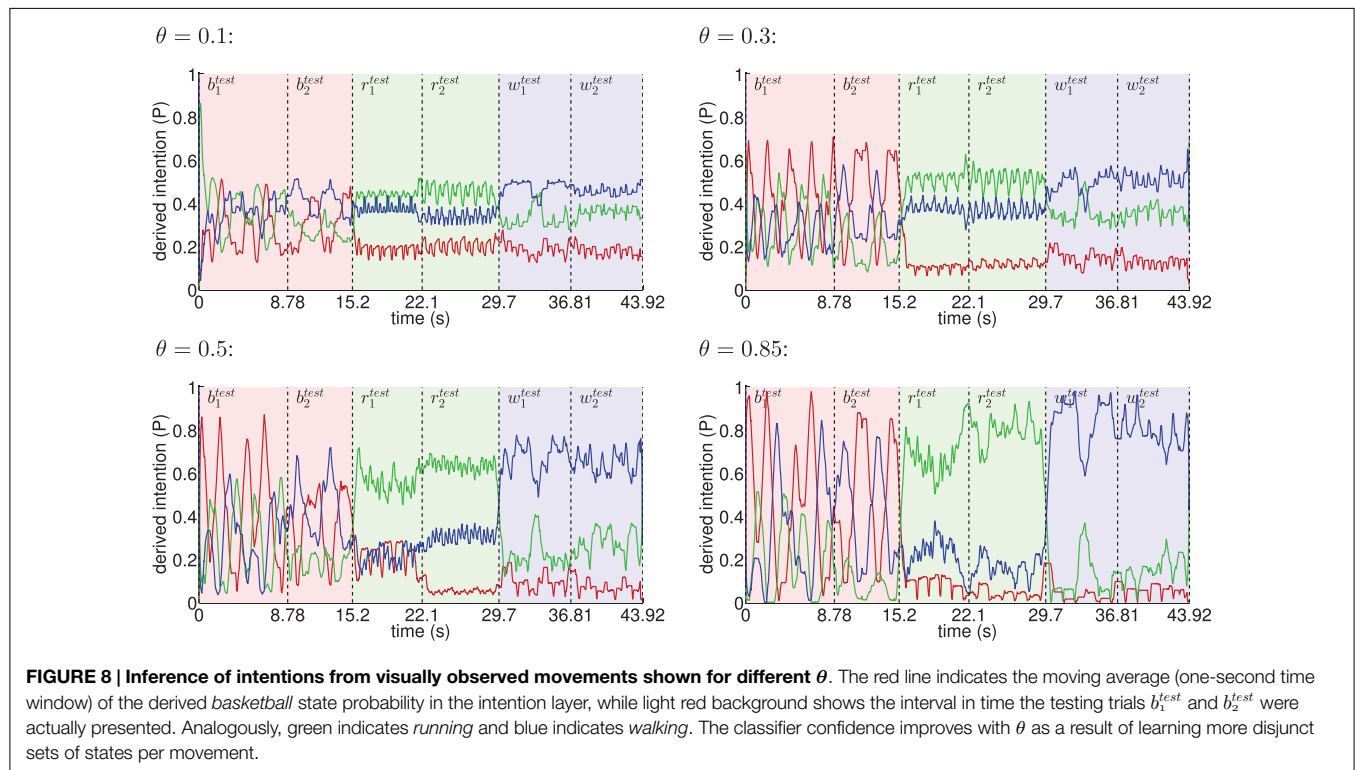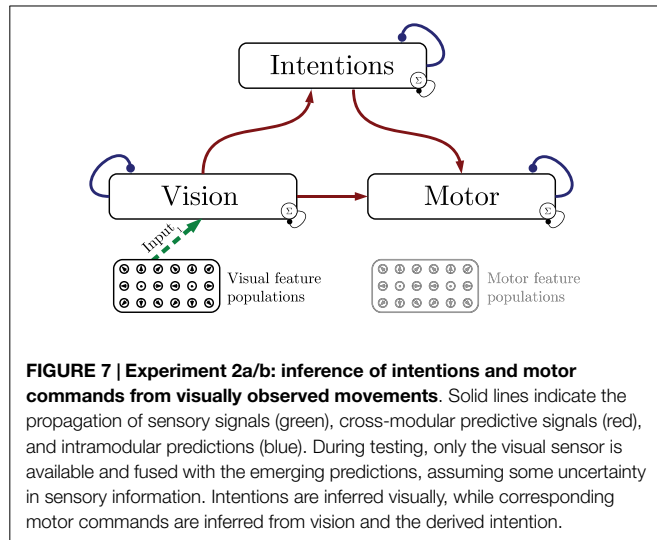
Classification results for four different θ averaged over 6 independent testing trials are shown in **Figure 8**. Despite the missing motor modality and the deviations in the observed posture control, the model was able to identify the character of the running and walking movements throughout, as concluded in **Table 1**. In doing so, accurately recognized visual state cells were enough to push the visual, motor, and intention state determination into temporal attractor sequences that consisted of the cyclic emulation of the respective movement using the embodied encodings. Following inputs then either maintained this emulation when close enough to the encodings or forced the convergence to

FIGURE 7 | Experiment 2a/b: inference of intentions and motor commands from visually observed movements. Solid lines indicate the propagation of sensory signals (green), cross-modular predictive signals (red), and intramodular predictions (blue). During testing, only the visual sensor is available and fused with the emerging predictions, assuming some uncertainty in sensory information. Intentions are inferred visually, while corresponding motor commands are inferred from vision and the derived intention.

another attractor sequence, that is, a shift in the perception. This effect can be seen clearly in the basketball trials, were episodes similar enough to the training data existed. However, as explained above, the basketball training trials were short and idealized, and they did not contain incongruent dribbling. The model then partly inferred a similarity with the trained walking movement in these segments, resulting in a bistable perception as shown in the graphs. This effect shows how the model is limited to the learned, embodied encodings when inferring intentions. It can be avoided by adding further training data.

When the learned movements were represented by a higher number of mainly disjunct states with respect to the movements, the model's ability to infer the intentions slightly improved. As a result of the more disjunct patterns, however, the confidence in classification improved consistently with θ from about 45 to 73% on average. As explained in the following, the classifier confidence has an influence on the inference of motor states.

5.3. Experiment 2b: Inference of Motor Commands upon Observation

Analogously to the preceding experiment, where we could show that intentions could be classified purely from visually observed motion patterns, we now evaluate if also the corresponding motor commands can be inferred using the same mechanisms. Potentially, this task is more difficult, since the set of available motor commands consists of a larger number of states in the motor layer when compared with the intention layer, and since the motor state transitions typically underlie faster dynamics. Seeing that the observed movements differed severely from the learned movements, we evaluate if the inferred motor state snapshots

FIGURE 8 | Inference of intentions from visually observed movements shown for different θ. The red line indicates the moving average (one-second time window) of the derived *basketball* state probability in the intention layer, while light red background shows the interval in time the testing trials b_1^{test} and b_2^{test} were actually presented. Analogously, green indicates *running* and blue indicates *walking*. The classifier confidence improves with θ as a result of learning more disjunct sets of states per movement.
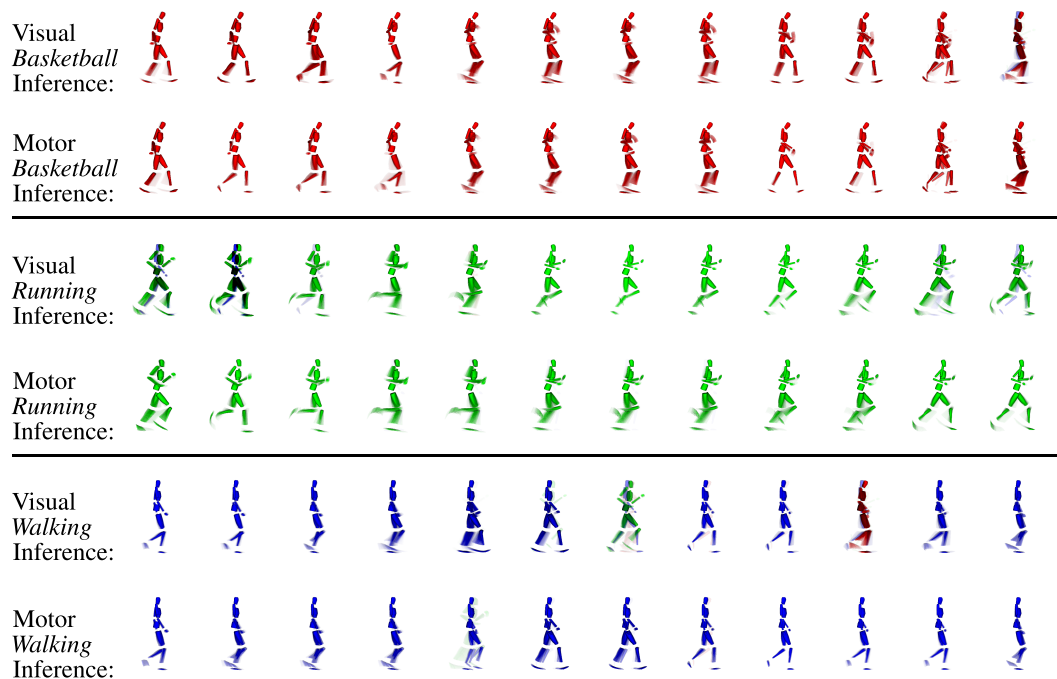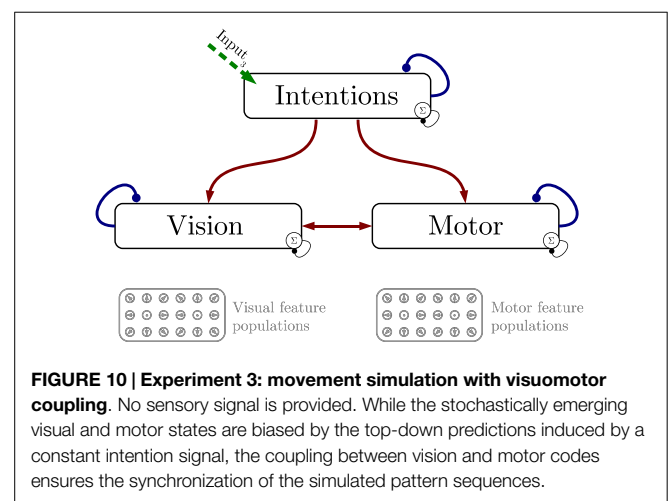
FIGURE 9 | Example clips of the state sequences recognized in the visual layer and inferred in the motor layer when observing three different movements (*basketball*, *running*, and *walking* testing trials). Each row shows the sequence of states by means of the representing snapshots over time (FLTR) for the respective modality and motion capture trial. Snapshots at the same position show the same time step in the sequence of visual and motor states and mostly show very similar parts of the movements. Because the inference is a stochastic process and because visual and motor states are not segmented in identical fashion as a result of the different information coding in the modalities, slight misalignments can occur. However, strong incongruence is avoided because of the visuomotor coupling. Moreover, although ambiguous patterns are included in the sequence, the network maintains the activation cycle of the movement-specific states because pattern transition probabilities are biased by top-down propagated intention signals.

correspond to the visual state snapshots at the same time steps and if the sequence in which they occur during the observation is plausible.

Figure 9 shows the coincidence of state snapshots of the recognized visual states and inferred motor states when observing the testing trials. When similar state snapshots are activated in both the visual and the motor domains, the two modalities can be considered to be synchronized in the emulation of the observed movement. In this process, both the cross-modular prediction from the vision to motor layer and the motor states predicted by the currently inferred intention bias the activation of motor states as indicated in **Figure 7**. The classifier confidence depicts the probability that a cell in the intention layer is selected as winner. Thus, increasing the classifier confidence will also increase the probability that movement-specific motor states are determined. Thus, since the classifier confidence increases with θ, the ability to imagine a sequence of motor codes corresponding to the currently observed visual motion, and the interpreted intention improves with the discretization of the state spaces.

5.4. Experiment 3: Simulation of Actions

Learning a tripartite model of visual motion states, corresponding motor codes, and intentions enables the inference of various bits of missing information. Seeing that information is encoded in normalized probability densities and information transfer is realized stochastically, activities in the network are self-sustaining even

FIGURE 10 | Experiment 3: movement simulation with visuomotor coupling. No sensory signal is provided. While the stochastically emerging visual and motor states are biased by the top-down predictions induced by a constant intention signal, the coupling between vision and motor codes ensures the synchronization of the simulated pattern sequences.

when sensory input is completely suppressed. When only provided with a top-down activation of a particular motion intention in the intention layer (cf. **Figure 10**), the model simulates likely sequences of modal visual and motor state sequences according to the learned temporal statistics.

In this experiment, we recorded the coinciding visual and motor state sequences generated by the model when a top-down intention-like action code is kept active in the intention layer. The results in **Figure 11** show that the learned sequences can
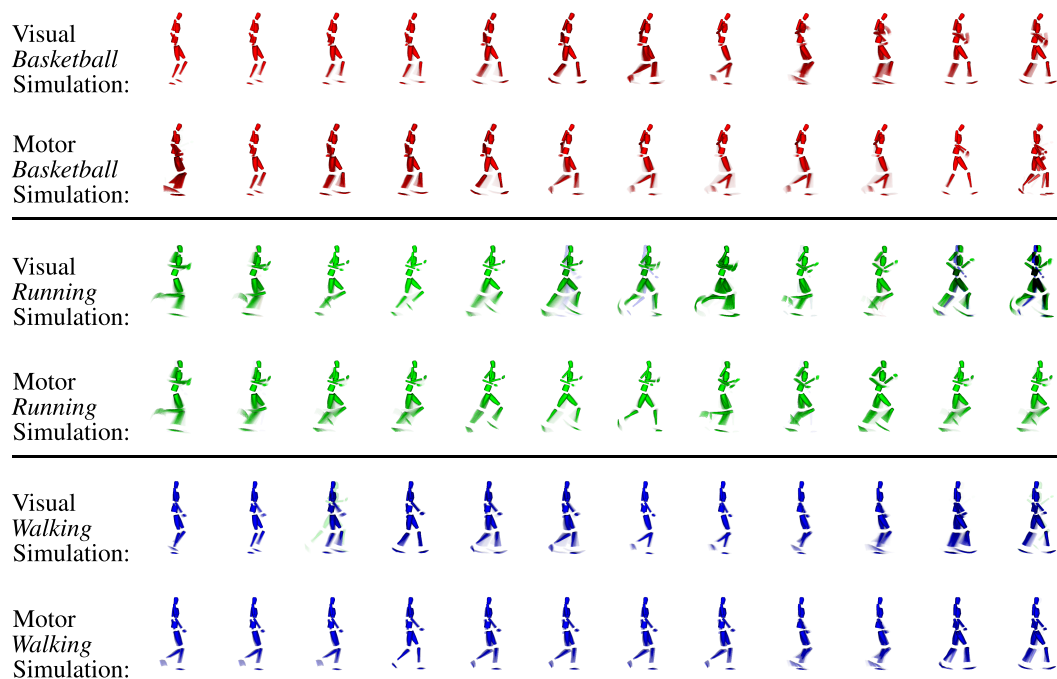
FIGURE 11 | Example clips of the state sequences simulated synchronously in the visual and motor modalities when three different intention priors (*basketball*, *running*, and *walking*) are provided. Each row shows the respective modality and intention prior. As during motor inference, snapshots at the same position mostly show very similar parts of the movements, while misalignments are mostly avoided. Again, the network maintains the activation cycle of the movement-specific states because pattern transition probabilities are biased by the provided top-down-propagated intention signals.

be replicated accurately both in the visual and in the motor domains. Although multiple ambiguous states were learned, as can be seen in the visual imaginations that are multi-colored, the simulated state sequence remains in the correct sequence and movement class. This is because the transition probabilities in the respective modalities are biased by the top-down intention signal.

The results also show that motor and visual state estimates remain approximately synchronized, seeing that the simulated states represent similar visual and motor imaginations at similar time steps. This indicates that the sensorimotor coupling is capable of synchronizing different modalities for periods of time. The reason for this synchronization lies in the lateral predictive connections between vision and motor layers: upon a transition from one visual state to another, the conditional probabilities for motor states given the new visual state change in an according fashion, such that the current motor state is more likely to transit to the most likely successor, which is not only determined by the top-down intention layer signal but also by the intramodular motor state transition probabilities and by the cross-modular activation predictions from the vision layer. Vice versa, the motor states bias the transition in the visual modality, leading to the observable mutual synchronization.

6. SUMMARY AND CONCLUSION

Our work shows that stochastic generative neural networks can be used to model action inference, mental imagery, and action simulation capabilities. Referring to Barsalou's simulation hypothesis,

it suggests that simulation processes in the brain may help to recognize, generalize, and maintain action perceptions and inferences using the own embodied encodings. In our model, these embodied simulations enable a consistent, multimodal interpretation of observed actions in abstract domains. In particular, we have shown that action observation models may rely on encodings that represent actions in a distributed and predictive manner: although some cells were encoding motion components that were active during the observation of various actions, cross-modular predictions enabled the consistent simulation of specific action sequences. Due to the predictive visuomotor coupling, temporal synchronicity of the activated states was ensured. Thus, the predictive, stochastic, and generative encodings resulted in the maintenance of overall consistent, multimodal motion imaginations. In combination with the previously published substructure of the model that resolves spatiotemporal variances by preprocessing of stimuli and inference of the perspective (Schrodt et al., 2015), a neural network architecture can be generated that infers the type of observed actions and possibly underlying motor commands, irrespective of the vantage point and despite variations of the movements. The model is thus able to establish the correspondence between self-perceptions and the perception of others, which can be considered an essential challenge in modeling action understanding.

Despite these successes, the model is currently based on several assumptions. For one, we assume that raw visual and motor perceptions and intentions can be simplified by compressed codes without losing model relevance, and that the respective motion features can be identified reliably. Although it is particularly

unclear how to incorporate realistic motor and intention codes in computational models, future model versions can be enhanced toward the processing of raw video streams of actions: the simulation snapshots in the experiments (see **Figures 9** and **11**) were calculated analogously to the conditional state predictions [equation (12)]. This shows that the states developed by the system can be suitably mapped onto lower level visual modalities. Thus, further developed models may hierarchically process lower level visual information similar to Jung et al. (2015), however, based on top-down predicted, higher level, and bodily grounded motion estimates.

Further, without sensory stimuli, the system's simulation of action states is a discrete time stochastic process. While the sequence of simulated states was mostly correct, the temporal duration of the activation was characterized by relatively high variance. Adding further modal state layers could diminish this variance. Particularly, the current model incorporates motion signals only and no static or postural information is processed. Exemplarily, the model implemented by Layher et al. (2014) triggers a reinforcement learning signal upon the encounter of low motion energy, which was used to foster the generation of posture snapshots in extreme poses. Comparably to the variance of simulated states, also the mean durations of state activations were partially distorted because of the approximate fusion of predicted state probability densities during testing. Integrating the systems predictions also during learning to a certain extent may improve the fusion of probabilities. It may also improve noise robustness and the establishment of disjunct modal state sequences. As shown in the experiments, disjunct states and state transitions are advantageous for the correct classification and emulation of actions. Techniques are available that can prevent the system to fall into an illusionary loop, when overly trusting the own predictions (Kneissler et al., 2014, 2015).

Moreover, the system currently simplifies a cell activation competition such that only one cell in each layer is adapted at each iteration. Using Mexican hat or softmax functions for the adaptation of learned states may speed up learning. Along similar lines, learning may be further improved when allowing a differential weighting of the provided input features. Currently, each input feature has the same influence in determining the creation of a new state. The recruitment of new prototypic states may be made dependent on the predictive value of all currently available states, including their specificity and accuracy, as is, for example, done in the XCSF learning classifier system architecture (Stalph et al., 2012; Kneissler et al., 2014). Another current

challenge to the system is to infer limb identities purely from visual information. The observed limb positions are fed into the dedicated neural network inputs. An adaptive confusion matrix could wire respective limb information appropriately, possibly by back-propagating mismatch signals. Additionally, lower level Gestalt constraints may be learned and used to adapt such a matrix.

Finally, despite the challenges remaining, also in its current form, the system may be evaluated as a cognitive model, and it may be used in robotics applications. Main predictions of the cognitive model come in the form of how visual motion will be segmented into individual motion clusters and how predictive encodings of the modalities modeled in the system will influence each other. Also, false information or distracting information from one module is expected to impair action recognition and simulation capabilities in the connected modules. On the robotics side, related techniques were applied using virtual visual servoing for object tracking (Comport et al., 2006) and for improving the pose estimates of a robot (Gratal et al., 2011). Our model offers both generative, visual servoing options and temporal motion predictions and inference-based, action recognition capabilities. In future work, this offers the opportunity to develop a cognitive system that is able to identify and subsequently emulate specific intention- or goal-oriented actions, striving for the same goal but adapting the motor commands to the own-bodily experiences and capabilities.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

REFERENCES

Baker, C. L., Saxe, R., and Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition* 113, 329–349. doi:10.1016/j.cognition.2009.07.005

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–600. doi:10.1017/S0140525X99532147

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi:10.1146/annurev.psych.59.103006.093639

Bonini, L., and Ferrari, P. F. (2011). Evolution of mirror systems: a simple mechanism for complex cognitive functions. *Ann. N. Y. Acad. Sci.* 1225, 166–175. doi:10.1111/j.1749-6632.2011.06002.x

Bruce, C., Desimone, R., and Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J. Neurophysiol.* 46, 369–384.

Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H.-J., et al. (2004). Neural circuits underlying imitation learning of hand actions: an event-related fMRI study. *Neuron* 42, 323–334. doi:10.1016/S0896-6273(04)00181-3

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., and Haggard, P. (2005). Action observation and acquired motor skills: an fMRI study with expert dancers. *Cereb. Cortex* 15, 1243–1249. doi:10.1093/cercor/bhi007

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., and Moore, C. (1998). Social cognition, joint attention, and communicative competence from

9 to 15 months of age. *Monogr. Soc. Res. Child Dev.* 63, i–vi, 1–143. doi:10.2307/1166214

Casile, A., Caggiano, V., and Ferrari, P. F. (2011). The mirror neuron system: a fresh view. *Neuroscientist* 17, 524–538. doi:10.1177/1073858410392239

Catmur, C., Walsh, V., and Heyes, C. (2007). Sensorimotor learning configures the human mirror system. *Curr. Biol.* 17, 1527–1531. doi:10.1016/j.cub.2007.08.006

Comport, A., Marchand, E., Pressigout, M., and Chaumette, F. (2006). Real-time markerless tracking for augmented reality: the virtual visual servoing framework. *IEEE Trans. Vis. Comput. Graph.* 12, 615–628. doi:10.1109/TVCG.2006.78

Cook, R., Bird, G., Catmur, C., Press, C., and Heyes, C. (2014). Mirror neurons: from origin to function. *Behav. Brain Sci.* 37, 177–192. doi:10.1017/S0140525X13000903

Dautenhahn, K., and Nehaniv, C. L. (2002). *The Correspondence Problem.* Massachusetts: MIT Press.

Elsner, B. (2007). Infants' imitation of goal-directed actions: the role of movements and action effects. *Acta Psychol.* 124, 44–59. doi:10.1016/j.actpsy.2006.09.006

Engel, A. K., Maye, A., Kurthen, M., and König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends Cogn. Sci.* 17, 202–209. doi:10.1016/j.tics.2013.03.006

Ferrari, P. F., Visalberghi, E., Paukner, A., Fogassi, L., Ruggiero, A., and Suomi, S. J. (2006). Neonatal imitation in rhesus macaques. *PLoS Biol.* 4:e302. doi:10.1371/journal.pbio.0040302

Friston, K., Mattout, J., and Kilner, J. (2011). Action understanding and active inference. *Biol. Cybern.* 104, 137–160. doi:10.1007/s00422-011-0424-z

Fritzke, B. (1995). A growing neural gas network learns topologies. *Adv. Neural Inf. Process Syst.* 7, 625–632.

Froese, T., Lenay, C., and Ikegami, T. (2012). Imitation by social interaction? Analysis of a minimal agent-based model of the correspondence problem. *Front. Hum. Neurosci.* 6:202. doi:10.3389/fnhum.2012.00202

Gallese, V. (2001). The 'shared manifold' hypothesis. From mirror neurons to empathy. *J. Conscious. Stud.* 8, 33–50.

Gallese, V. (2007a). Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 659–669. doi:10.1098/rstb.2006.2002

Gallese, V. (2007b). Embodied simulation: from mirror neuron systems to interpersonal relations. *Novartis Found. Symp.* 278, 3–12. doi:10.1002/9780470030585.ch2

Gallese, V., and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* 2, 493–501. doi:10.1016/S1364-6613(98)01262-5

Gallese, V., Rochat, M., Cossu, G., and Sinigaglia, C. (2009). Motor cognition and its role in the phylogeny and ontogeny of action understanding. *Dev. Psychol.* 45, 103. doi:10.1037/a0014436

Garcia, J. O., and Grossman, E. D. (2008). Necessary but not sufficient: motion perception is required for perceiving biological motion. *Vision Res.* 48, 1144–1149. doi:10.1016/j.visres.2008.01.027

Gergely, G., Nádasdy, Z., Csibra, G., and Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition* 56, 165–193. doi:10.1016/0010-0277(95)00661-H

Giese, M. A., and Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nat. Rev. Neurosci.* 4, 179–192. doi:10.1038/nrn1057

Gratal, X., Romero, J., and Kragic, D. (2011). "Virtual visual servoing for real-time robot pose estimation," in *World Congress*, Vol. 18, eds S. Bittanti, A. Cenedese, and S. Zampieri (Milano: International Federation of Automatic Control), 9017–9022.

Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Stud. Appl. Math.* 52, 213–257. doi:10.1002/sapm1973523213

Grossberg, S. (1976a). On the development of feature detectors in the visual cortex with applications to learning and reaction-diffusion systems. *Biol. Cybern.* 21, 145–159. doi:10.1007/BF00337422

Grossberg, S. (1976b). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biol. Cybern.* 23, 121–134. doi:10.1007/BF00344744

Grossberg, S. (1976c). Adaptive pattern classification and universal recoding: II. Feedback, expectation, olfaction, illusions. *Biol. Cybern.* 23, 187–202.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., et al. (2000). Brain areas involved in perception of biological motion. *J. Cogn. Neurosci.* 12, 711–720. doi:10.1162/089892900562417

Heyes, C. (2001). Causes and consequences of imitation. *Trends Cogn. Sci.* 5, 253–261. doi:10.1016/S1364-6613(00)01661-2

Heyes, C. (2010). Where do mirror neurons come from? *Neurosci. Biobehav. Rev.* 34, 575–583. doi:10.1016/j.neubiorev.2009.11.007

Iacoboni, M. (2005). Neural mechanisms of imitation. *Curr. Opin. Neurobiol.* 15, 632–637. doi:10.1016/j.conb.2005.10.010

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annu. Rev. Psychol.* 60, 653–670. doi:10.1146/annurev.psych.60.110707.163604

Iacoboni, M., and Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nat. Rev. Neurosci.* 7, 942–951. doi:10.1038/nrn2024

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* 14, 201–211. doi:10.3758/BF03212378

Jung, M., Hwang, J., and Tani, J. (2015). Self-organization of spatio-temporal hierarchy via learning of dynamic visual image patterns on action sequences. *PLoS ONE* 10:e0131214. doi:10.1371/journal.pone.0131214

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends Cogn. Sci.* 15, 352–357. doi:10.1016/j.tics.2011.06.005

Kilner, J. M., Friston, K. J., and Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cogn. Process.* 8, 159–166. doi:10.1007/s10339-007-0170-2

Kneissler, J., Drugowitsch, J., Friston, K., and Butz, M. V. (2015). Simultaneous learning and filtering without delusions: a bayes-optimal combination of predictive inference and adaptive filtering. *Front. Comput. Neurosci.* 9:47. doi:10.3389/fncom.2015.00047

Kneissler, J., Stalph, P. O., Drugowitsch, J., and Butz, M. V. (2014). Filtering sensory information with XCSF: improving learning robustness and robot arm control performance. *Evol. Comput.* 22, 139–158. doi:10.1162/EVCO_a_00108

Lallee, S., and Dominey, P. F. (2013). Multi-modal convergence maps: from body schema and self-representation to mental imagery. *Adapt. Behav.* 21, 274–285. doi:10.1177/1059712313488423

Layher, G., Giese, M. A., and Neumann, H. (2014). Learning representations of animated motion sequences – A neural model. *Top. Cogn. Sci.* 6, 170–182. doi:10.1111/tops.12075

Lepage, J.-F., and Théoret, H. (2007). The mirror neuron system: grasping others' actions from birth? *Dev. Sci.* 10, 513–523. doi:10.1111/j.1467-7687.2007.00631.x

Meltzoff, A. N. (2007). 'Like me': a foundation for social cognition. *Dev. Sci.* 10, 126–134. doi:10.1111/j.1467-7687.2007.00574.x

Nagai, Y., Kawai, Y., and Asada, M. (2011). "Emergence of mirror neuron system: immature vision leads to self-other correspondence," in *2011 IEEE International Conference on Development and Learning (ICDL)*, Vol. 2 (Frankfurt am Main: IEEE), 1–6.

Oja, E. (1989). Neural networks, principal components, and subspaces. *Int. J. Neural Syst.* 1, 61–68. doi:10.1142/S0129065789000475

Oram, M., and Perrett, D. I. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to "biological motion" stimuli. *J. Cogn. Neurosci.* 6, 99–116. doi:10.1162/jocn.1994.6.2.99

Pavlova, M. A. (2012). Biological motion processing as a hallmark of social cognition. *Cereb. Cortex* 22, 981–995. doi:10.1093/cercor/bhr156

Perrett, D., Smith, P., Mistlin, A., Chitty, A., Head, A., Potter, D., et al. (1985). Visual analysis of body movements by neurons in the temporal cortex of the macaque monkey: a preliminary report. *Behav. Brain Res.* 16, 153–170. doi:10.1016/0166-4328(85)90089-0

Pouget, A., Dayan, P., and Zemel, R. (2000). Information processing with population codes. *Nat. Rev. Neurosci.* 1, 125–132. doi:10.1038/35039062

Puce, A., and Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 435–445. doi:10.1098/rstb.2002.1221

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

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192. doi:10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., and Craighero, L. (2005). "Mirror neuron: a neurological approach to empathy," in *Neurobiology of Human Values*, eds J.-P. Changeux, A. R. Damasio, W. Singer, and Y. Christen (Heidelberg: Springer), 107–123.

Saby, J. N., Marshall, P. J., and Meltzoff, A. N. (2012). Neural correlates of being imitated: an eeg study in preverbal infants. *Soc. Neurosci.* 7, 650–661. doi:10.1080/17470919.2012.691429

Schrodt, F., and Butz, M. V. (2014). "Modeling perspective-taking by forecasting 3D biological motion sequences," in *Cognitive Processing, Suppl. KogWis 2014*, Vol. 15, eds M. O. Belardinelli, A. Belardinelli, and M. V. Butz (Tübingen: Springer), 137–139.

Schrodt, F., Layher, G., Neumann, H., and Butz, M. V. (2014a). "Modeling perspective-taking by correlating visual and proprioceptive dynamics," in *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (Quebec City), 1383–1388.

Schrodt, F., Layher, G., Neumann, H., and Butz, M. V. (2014b). "Modeling perspective-taking upon observation of 3D biological motion," in *Proceedings of the 4th International Conference on Development and Learning and on Epigenetic Robotics* (Genoa), 328–333.

Schrodt, F., Layher, G., Neumann, H., and Butz, M. V. (2015). Embodied learning of a generative neural model for biological motion perception and inference. *Front. Comput. Neurosci.* 9:79. doi:10.3389/fncom.2015.00079

Stalph, P. O., Llorá, X., Goldberg, D. E., and Butz, M. V. (2012). Resource management and scalability of the xcsf learning classifier system. *Theor. Comp. Sci.* 425, 126–141. doi:10.1016/j.tcs.2010.07.007

Taylor, G. W., Hinton, G. E., and Roweis, S. T. (2006). "Modeling human motion using binary latent variables," in *Advances in Neural Information Processing Systems 19*, eds S. Bernhard, P. John, and H. Thomas (MIT Press), 1345–1352.

Thurman, S. M., and Grossman, E. D. (2008). Temporal 'bubbles' reveal key features for point-light biological motion perception. *J. Vis.* 8, 1–11. doi:10.1167/8.3.28

Tomasello, M. (1999). The human adaptation for culture. *Annu. Rev. Anthropol.* 28, 509–529. doi:10.1146/annurev.anthro.28.1.509

Turella, L., Wurm, M. F., Tucciarelli, R., and Lingnau, A. (2013). Expertise in action observation: recent neuroimaging findings and future perspectives. *Front. Hum. Neurosci.* 7:637. doi:10.3389/fnhum.2013.00637

Ulloa, E. R., and Pineda, J. A. (2007). Recognition of point-light biological motion: mu rhythms and mirror neuron activity. *Behav. Brain Res.* 183, 188–194. doi:10.1016/j.bbr.2007.06.007

Want, S. C., and Harris, P. L. (2002). How do children ape? Applying concepts from the study of non-human primates to the developmental study of 'imitation' in children. *Dev. Sci.* 5, 1–14. doi:10.1111/1467-7687.00194

CrossMark

Exploration Behaviors, Body Representations, and Simulation Processes for the Development of Cognition in Artificial Agents

Guido Schillaci[1]*, Verena V. Hafner[1] and Bruno Lara[2]

[1] *Adaptive Systems Group, Department of Computer Science, Humboldt-Universität zu Berlin, Berlin, Germany,* [2] *Cognitive Robotics Group, Center for Science Research, Universidad Autonoma del Estado de Morelos, Cuernavaca, Mexico*

Sensorimotor control and learning are fundamental prerequisites for cognitive development in humans and animals. Evidence from behavioral sciences and neuroscience suggests that motor and brain development are strongly intertwined with the experiential process of *exploration*, where internal body representations are formed and maintained over time. In order to guide our movements, our brain must hold an internal model of our body and constantly monitor its configuration state. How can sensorimotor control enable the development of more complex cognitive and motor capabilities? Although a clear answer has still not been found for this question, several studies suggest that processes of mental simulation of action–perception loops are likely to be executed in our brain and are dependent on internal body representations. Therefore, the capability to re-enact sensorimotor experience might represent a key mechanism behind the implementation of higher cognitive capabilities, such as behavior recognition, arbitration and imitation, sense of agency, and self–other distinction. This work is mainly addressed to researchers in autonomous motor and mental development for artificial agents. In particular, it aims at gathering the latest developments in the studies on exploration behaviors, internal body representations, and processes of sensorimotor simulations. Relevant studies in human and animal sciences are discussed and a parallel to similar investigations in robotics is presented.

Keywords: sensorimotor learning, exploration behaviors, body representations, internal models, sensorimotor simulations, developmental robotics

1. INTRODUCTION

The capability to perform sensory-guided motor behaviors, or sensorimotor control, is generally not fully developed at birth in mammals. Rather, it emerges through a learning process where the individual is actively involved in the interaction with the external environment. In humans, similarly, sensorimotor control is developed along the ontogenetic process of the individual. Developmental psychologists consider this skill as a fundamental prerequisite for the acquisition of more complex cognitive and social capabilities.

In robotics, a large number of studies investigated mechanisms for sensorimotor control and learning in artificial agents. Inspired by human development and aiming at producing adaptive systems (Asada et al., 2009; Law et al., 2011), researchers proposed robot learning mechanisms based

on exploration behaviors. Evidence from human behavioral and brain sciences suggests that motor and brain development are strongly intertwined with this experiential process, where internal body representations would be formed and maintained over time. However, it is still not clear how sensorimotor development is linked to the development of cognitive skills. Indeed, one of the most challenging questions in developmental sciences, including developmental robotics, is how low-level motor skills scale up to more complex motor and cognitive capabilities throughout the lifespan of an individual.

A recent line of thought identifies the capability to internally simulate sensorimotor cycles based on previous experience, or to *re-enact* past sensorimotor experience, as one of the fundamental processes implicated in the implementation of cognitive skills (Barsalou, 2008). Several behavioral and brain studies can be found in the literature that support this idea. In this work, we argue that sensorimotor simulation mechanisms may serve as a bridge between sensorimotor representation and the implementation of basic cognitive skills, such as behavior recognition, arbitration and imitation, sense of agency, and self–other distinction. During the last years, an increasing number of robotics studies addressed similar processes for the implementation of cognitive skills in artificial agents. However, empirical investigation on exploration behaviors for the learning of sensorimotor control, on the functioning and modeling of simulation processes in the brain, and on their implementation in artificial agents is still fragmented.

This paper aims at gathering the latest developments in the study on exploration behaviors, or internal body representations, and on re-using sensorimotor experience for cognition. For each of these topics, relevant studies in human and animal sciences will be introduced and similar studies in robotics will be discussed. We strongly believe that this can be beneficial for those researchers who investigate autonomous motor and mental development for artificial agents. This manuscript provides a comprehensive overview of the state of the art in the mentioned topics from different perspectives. Moreover, we want to encourage robotic researchers in sensorimotor learning and in body representations to make a step further by investigating how the acquired sensorimotor experience can be used for cognition. Nonetheless, we would like to encourage researchers to not overlook the process of acquisition of sensorimotor experience by assuming the existence of a repertoire of sensorimotor schemes, when investigating computational models for internal simulations. We strongly believe that tackling both issues at the same time not only would provide a more comprehensive view of the developmental process in artificial agents but it would also give insights into the generalization and specialization of the proposed models. In addition, we believe that addressing both sensorimotor and cognitive development by simulation processes would bridge different specialties and provide new research directions for developmental robotics.

2. EXPLORATION AS A DRIVE FOR MOTOR AND COGNITIVE DEVELOPMENT

In the late 1980s, a new era known as post-cognitivism started to flourish in the cognitive sciences, bringing new philosophical interest on embodiment and on the importance of the role of the body for cognition (Wilson and Foglia, 2011). According to the embodied cognition framework, sensorimotor interaction is essential for the development of cognition. A common characteristic of humans, animals, and artificial agents is their embodiment and their being situated in an environment they can interact with. They possess the means for shaping these interactions: a body that can be actuated by controlling its muscles (in humans and animals) or actuators (in artificial agents) and the capability to perceive internal or external phenomena through their senses (in humans and animals) or sensors (in artificial agents) (Pfeifer and Bongard, 2006). In animals and humans, brain development is modulated by the multimodal sensorimotor information experienced by the individual while interacting with the external environment. In the literature, this process is often referred as *sensorimotor learning*. Theorists on grounded cognition propose that cognitive capabilities are grounded on sensorimotor experiences (Barsalou, 2008). Although the validity of this theory is still under debate, it is commonly accepted that sensorimotor control and learning are fundamental prerequisites for cognitive development in humans. Therefore, developmental roboticists are particularly interested in implementing exploration behaviors in artificial agents, which would allow them to gather the necessary sensorimotor experience to further develop complex motor and cognitive skills.

Humans are not innately skillful at governing their body. Motor control is a capability that is acquired and refined over time, as demonstrated by several studies. For example, Zoia and colleagues have shown that learning of motor control is an ongoing process already during pre-natal stages (Zoia et al., 2007). In fact, they observed an improvement of coordinated kinematic patterns in fetuses between the age of 18 and 22 weeks. At initial stages, fetuses' hand movements directed at their eyes and mouth were inaccurate and characterized by jerky and zigzag movements. However, already around the 22nd week of gestation, fetuses showed more precise hand trajectories, characterized by acceleration and deceleration phases that were apparently planned according to the size and to the delicacy of the target (facial parts, such as mouth or eyes). It is plausible to think that such an improvement in sensorimotor control would be the result of an experiential process, driven by exploration behaviors. Many developmental studies agree with this hypothesis (Piaget, 1954; Kuhl and Meltzoff, 1996; Thelen and Smith, 1996; Meltzoff and Moore, 1997). Others show systematic exploration behaviors already at early stages of post-natal development [for example, in the visual and proprioceptive domains (Rochat, 1998)].

In developmental psychology, exploration behaviors are seen as the common characteristic of initial stages of motor and cognitive development. In an early study, Jean Piaget defined exploration behaviors as *circular reactions* – or repetitions of movements that the child finds pleasurable – through which infants gather experience and acquire governance of those motor capabilities (such as reaching an object) that will enable them, subsequently, to explore the interactions with objects and with people (Piaget, 1954). Therefore, exploration would pave the way to the development of more complex motor and social capabilities. In a study on language acquisition, for example, Kuhl and Meltzoff (1996) reported that in infants younger than 6 months the vocal tract and

the neuromusculature are still immature for the production of recognizable sounds. It is through exploratory behaviors, which Meltzoff and Moore (1997) named as *body* or *vocal babbling*, that infants would learn articulatory–auditory relations, a prerequisite for language acquisition.

However, it is not clear what the drive of exploration behaviors is. Behavioral studies agree with the fact that animals and humans seem to have a common desire to experience and to acquire new information (Berlyne, 1960; Reio et al., 2006; Reio, 2011). Such a characteristic, commonly referred to as *curiosity*, is usually associated with the experience of rewards, similar to appetitive desires for food and sex (Litman, 2005). However, several theories have been developed on the mechanisms that explain curiosity. For example, the curiosity-driven theory assumes that organisms are motivated to acquire new information through exploratory behaviors by the need of restoring cognitive and perceptual coherence (Berlyne, 1960). Such a coherence can be disrupted by an unpleasant experience of uncertainty, an unpleasant feeling of deprivation, the reduction of which is rewarding (Litman, 2005).

Curiosity and exploration behaviors are considered as fundamental aspects of learning and development. However, studying them in humans and animals often means to observe and to analyze only their behavioral effects, which is of course limiting the understanding of the underlying processes. Robots recently came into play as they provide a valuable test bed for the investigation of such mechanisms. Investigating curiosity and exploration behaviors in artificial agents is indeed also advantageous for developmental roboticists, whose aim is to produce autonomous, adaptive, and social robots, which learn from and adapt to the dynamic environment using mechanisms inspired by human development (Lungarella et al., 2003). The developmental approach in robotics is not only motivated by a mere interest in mimicking human development in artificial agents. Rather, studying human development can give insights in finding those basic behavioral components that may allow for the autonomous mental and motor development in artificial agents. In fact, researchers in developmental robotics try to avoid defining models of robot embodiment and of their surrounding world *a priori*, in order to not stumble across problems, such as robot behaviors lacking adaptability and the capability to react to unexpected events (Schillaci, 2014).

In developmental robotics, the general approach consists of providing artificial agents with learning mechanisms based on exploration behaviors. In addition to humans, robots can generate useful information about their bodily capabilities while interacting with the external environment. This information is shaped by the characteristics of the agent's body and of the environment. In addition, dynamic environments and temporary or permanent changes in the bodily characteristics of the individual, for example, the ones caused by the usage of tools, can strongly affect the information that is perceived through the senses and the way the individual can interact with its surroundings. Therefore, pre-defining models of the robot's body and of the environment can be very challenging, or even impossible, as an enormous number of variables have to be taken into account, for covering all the aspects of such dynamic systems. This is one of the main motivations behind the developmental approach in robotics,

where researchers try to implement computational models that self-organize along the sensorimotor information that is generated from the bodily interaction of the agent with the external environment, such as the one produced through exploration behaviors, while assuming as little prior information to construct the model as possible.

Several studies on the development of motor and cognitive skills based on exploration behaviors can be found in the literature on developmental robotics. For example, in a survey on cognitive developmental robotics, Asada and colleagues presented a developmental model of human cognitive functions starting from the fetal simulation of sensorimotor learning of body representation in the womb up to the social development through interaction between individuals, namely imitation (Asada et al., 2009). The authors put a central role to exploration behaviors for the emergence of cognition in infants and artificial agents. These behaviors are the drive to the construction of body representations (see Section 3), or mappings of multimodal sensorimotor information, which are necessary for interacting with the external environment, for example, with objects. Learning of coordinated movements, such as reaching and grasping, is considered to develop along the infant's acquisition of predictive capabilities, which may play an important role in the development of non-verbal communication, such as pointing or imitation (Asada et al., 2009; Hafner and Schillaci, 2011).

Dearden and Demiris (2005) also adopted exploration behaviors for learning internal forward models in an artificial agent. As it will be discussed in more details in the following sections of this paper, forward models enable a robot to predict the consequence of its motor actions. In Dearden and Demiris (2005), a robot performed random movements of its gripper and visually observed the outcome of these actions. The internal forward model was encoded as a Bayesian network, whose structure and parameters were learned using the sensorimotor data gathered during the exploration behavior performed in the motor space. This exploration strategy, known also as random body babbling, chooses motor commands from the range of possible movements in a random fashion. Takahashi and colleagues implemented a similar exploration mechanism in a simulated robotic setup for learning motion primitives under tool-use conditions (Takahashi et al., 2014). A simulated robotic arm was programed to execute an exploration behavior – random body babbling – in order to gather sensorimotor information to be used for building up a body representation. The authors adopted a recurrent neural network for training the body representation and a deep neural network for encoding the tool dynamic features and evaluated the approach in an object manipulation task.

Stoytchev (2005) presented an experiment with a simulated robot on learning the binding affordances of objects using pre-defined exploration motion primitives that were selected in a random fashion. The action opportunities that an object provides to the agent, or affordances, were learned during the exploration session where the robot randomly chose sequences of pre-defined behaviors, applied them to explore the objects, and detected invariants in the resulting set of observations. However, the proposed approach is limited by the usage of pre-defined movements and by the lack of variability in the exploration behaviors. In fact,

there could be object affordances that are unlikely to be discovered due to the unavailability of specific exploratory behaviors.

Many other developmental robotics studies adopting random exploration behaviors, and comparing different random movement strategies (Schillaci and Hafner, 2011), can be found in the literature. However, random exploration strategies – such as random body babbling, or motor babbling (see **Figure 1**) – have been found to be not optimal, especially when applied to robotic systems characterized by a high number of degrees of freedom. More efficient sensorimotor exploration behaviors have been proposed. For example, Baranes and Oudeyer (2013) presented an intrinsically motivated goal exploration mechanism that allows active learning of inverse models, or controllers, in redundant robots. In the proposed methodology, exploration is performed in the task space, making it more efficient than exploring the motor space, especially when using high-dimensional robots. Rolf and Steil showed a similar approach based on goal-directed exploration in the task space, which enabled successful learning of the controller on a challenging robot platform, the Bionic Handling Assistant (Rolf and Steil, 2014).

Investigating goal-directed exploration behaviors provided new insights and research directions toward the understanding of the mechanisms behind curiosity. In fact, one of the main questions posed by researchers on goal-directed exploration in artificial agents is how to generate goals in the task space. The typical approach proposes to simulate curiosity in an artificial agent, by adding interest factors in the exploration phase, usually based on measuring the confidence that the system has toward possible goals in the space to be explored. Information seeking through exploration behaviors, according to Gottlieb et al. (2013), is "a process that obeys the imperative to reduce uncertainty and can be extrinsically or intrinsically motivated." This is in line with what has been proposed by Litman (2005), as mentioned before in this section, that the drive of curiosity might rely on the reduction of the *unpleasant* experience of uncertainty, which is rewarding. Oudeyer et al. (2007), Baranes and Oudeyer (2013), Moulin-Frier et al. (2013), and Schmerling et al. (2015) adopted an intrinsically motivated goal exploration mechanism, named Intelligent

Adaptive Curiosity (IAC), which relies on the uncertainty reduction idea and on exploration based on learning progress. In other words, IAC selects goals maximizing a competence progress, thus creating developmental trajectories driving the robot to progressively focus on tasks of increasing complexity and is statistically significantly more efficient than selecting tasks in a random fashion. IAC has been applied to different contexts in artificial agents, such as in learning sensorimotor affordances (Oudeyer et al., 2007), in learning inverse kinematics of a simulated robotic arm and in learning motor primitives in mobile robots (Baranes and Oudeyer, 2013), in vocal learning (Moulin-Frier et al., 2013), in the context of oculomotor coordination (Gottlieb et al., 2013), and in learning visuo-motor coordination in a humanoid robot (Schmerling et al., 2015). In this latter study, in particular, the authors showed not only the superiority of goal-directed exploration strategies, compared to random ones, but also their effectiveness in the case where two separate motor sub-systems, head and arm in the presented experiment, need to be coordinated.

Other experiments on curiosity-driven exploration behaviors can be found in the literature. However, most of them adopt an approach similar to IAC that implements exploration mechanisms based on the learning progress. Ngo et al. (2013), for example, proposed a system that generates goals based on the confidence in its predictions about how the environment reacts to its actions; when the confidence on a prediction is low, the environmental configuration that generated such an event becomes a goal. Pape et al. (2012) presented a similar curiosity-driven exploration behavior in the context of tactile skills learning, which allowed the robotic system to autonomously develop a small set of basic motor skills that lead to different kinds of tactile input, and to learn how to exploit the learned motor skills to solve texture classification tasks. Jauffret et al. (2013) presented a neural architecture based on an online novelty detection algorithm that is able to self-evaluate sensory-motor strategies. Similar to the abovementioned mechanism, in the proposed system, the prediction error coming from unexpected events provides a measure of the quality of the underlying sensory-motor schemes and it is used to modulate the system's behavior
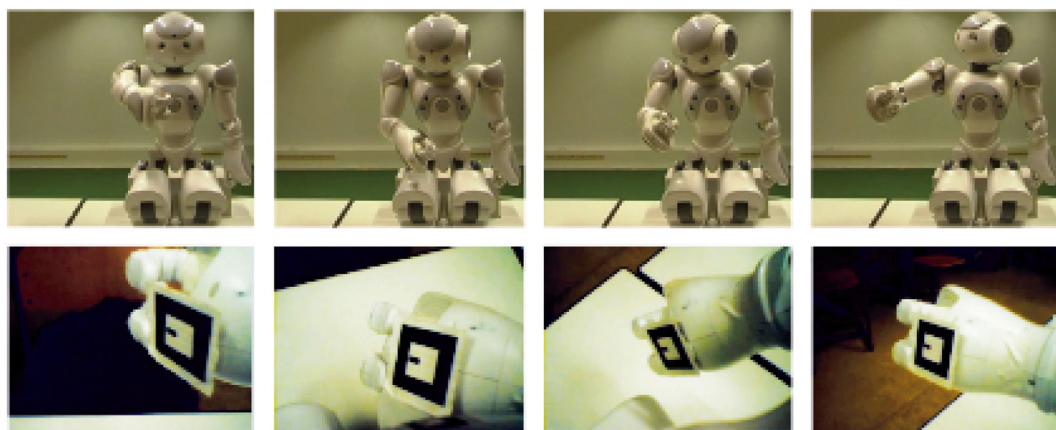

FIGURE 1 | A sequence from an exploration behavior – in this case random motor babbling – performed by the humanoid robot Aldebaran Nao. In the bottom, the corresponding frames grabbed by the robot camera are shown. Picture taken from Schillaci and Hafner (2011).

in a navigation task. A greedy goal-directed exploration strategy has been adopted, instead, by Berthold and Hafner (2015), who presented an approach for online learning of a controller for a low-dimensional spherical robot based on reservoir computing. The exploration strategy adopted by the authors generated motor commands aimed at regulating the sensory input to externally generated target values.

The number of robotics studies investigating sensorimotor exploration behaviors for robot learning has been considerably growing in the last couple of decades. This section mentioned the most prominent studies, with a particular focus on the competences that such exploration behaviors allowed the robots to acquire. **Table 1** summarizes the studies that are cited in this work, and for each of them it points out whether and what exploration strategies have been used for learning particular skills. As evident in these descriptions, most of the studies addressing intelligent exploration behaviors, or exploratory strategies that try to mimic human curiosity, are prevalently adopted only for

TABLE 1 | Summary of robotic studies discussed in this work.

Robotic study	Exploration behaviors	Body representation	Predictive processes	Skills implemented	Simulated or real robot	Learning tool
Cognitive developmental robotics: a survey (Asada et al., 2009)	Yes	Yes	Yes	Spatial perception, vocal imitation	Fetal simulation and real humanoid robot	Several tools, including self-organizing maps and Hebbian learning
Learning forward models for robots (Dearden and Demiris, 2005)	Random motor babbling	Not explicitly addressed	Yes	Recognition and imitation of hand waving movements	Robot gripper and camera	Bayesian network
Toward learning the binding affordances of objects: a behavior-grounded approach (Stoytchev, 2005)	Random selection of pre-defined behaviors	Not explicitly addressed	No	Learning of object affordances	Simulated robotic arm	Search in history table storing behaviors/ affordance mappings
Tool-body assimilation model based on body babbling and neurodynamical system (Takahashi et al., 2014)	Random body babbling	Yes	No	Object manipulation using tools	Simulated robotic arm and tools	Recurrent neural network and deep neural network
Oudeyer et al. (2007), Baranes and Oudeyer (2013), Moulin-Frier et al. (2013), and Schmerling et al. (2015)	Intrinsically motivated (goal-directed) exploration	Sensorimotor mappings	Yes, during the exploration	Visuo-motor coordination, learning of sensorimotor affordances, motion primitives, and model-based vocal production	Simulated and real robots	k-NN-based search, Gaussian mixture models
Learning tactile skills through curious exploration (Pape et al., 2012)	Curiosity-driven reinforcement learning (goal-directed exploration)	Not explicitly addressed	Yes, during the exploration	Tactile exploratory and texture classification skills	Biomimetic robotic finger	Reinforcement learning
Babyrobot – a study on sensori-motor development (Metta, 2000)	Goal-directed exploration	Self-organizing maps and multimodal mapping	Yes	Visuo-motor coordination	Robot head and manipulator	Self-organizing neural networks
Active motor babbling for sensorimotor learning (Saegusa et al., 2008)	Exploration based on prediction error dependent confidence measures	Visuo-motor mapping	Yes, in estimating the prediction error dependent confidence	Visuo-motor coordination	Humanoid robot	Multi-layer perceptrons
Epigenetic Robotics Architecture (Morse et al., 2010)	Not explicitly addressed	Self-organizing maps for multimodal mappings	Spreading of activations between maps	Grounding of linguistic labels onto body postures, visual and auditory modalities	Humanoid robot	Self-organizing maps and Hebbian learning
Kajić et al. (2014) and Schillaci et al. (2014)	Random motor exploration	Self-organizing maps for sensorimotor mappings	Yes, for executing reaching and pointing behaviors	Visuo-motor coordination and proto-imperative pointing gestures	Humanoid robot	Self-organizing maps and Hebbian learning
Schillaci et al. (2012a) and Schillaci (2014)	Random motor exploration	Sensorimotor mappings encoded in internal models	Internal simulations based on inverse and forward predictions	Action selection for tool-use	Humanoid robot	k-NN and multi-layer perceptrons

(Continued)

TABLE 1 | Continued

Robotic study	Exploration behaviors	Body representation	Predictive processes	Skills implemented	Simulated or real robot	Learning tool
Reaching for the unreachable: integration of locomotion and whole-body movements for extended visually guided reaching (Brandao et al., 2013)	Goal-directed reaching movements	Kinematic models	Yes, in the reaching behavior	Gaze-centered whole-body reaching and locomotion	Humanoid robot	Non-linear regression (LWPR)
Representations of body schemas for infant robot development (Shaw et al., 2015)	Intrinsic motivation-based exploration	Yes, sensorimotor maps	Yes, in the exploration behavior	Staged development of eye saccades, gaze control, primitive reaching, coordinated reach-to-seen and gaze-to-touch behaviors	Humanoid robot	Topological maps and Hebbian learning
Interpersonal maps: how to map affordances for interaction behavior (Hafner and Kaplan, 2008)	Precoded motion primitives	Body maps and interpersonal maps based on information distance	Yes, in imitation mechanisms	Imitation between agents	Robot dogs	Distance between information sources
Ince et al. (2009)	Precoded motion primitives	Mapping between motor commands and auditory ego-noise	Predictions for ego-noise suppression	Ego-noise modeling and suppression	Humanoid robot	k-NN in a ego-noise template database
Internal simulation of perception: a minimal neuro-robotic model (Ziemke et al., 2005)	Navigation with an evolved controller	Not explicitly address	Yes	Prediction of non-present sensory input for robot navigation	Simulated mobile robot	Recurrent neural networks and evolution
Self body mapping in mobile robots using vision and forward models (Escobar et al., 2012)	Random exploration of the environment	Not explicitly addressed	Yes	Long-term predictions, collision-free navigation	Simulated and real mobile robot	Multi-layer perceptrons
Action selection and mental transformation based on a chain of forward models (Hoffmann and Möller, 2004)	Random exploration of the environment	Not explicitly addressed	Yes	Long-term predictions for robot navigation	Real mobile robot	Multi-layer perceptrons
Contingency perception and agency measure in visuo-motor spiking neural networks (Pitti et al., 2009)	Random exploration	Not explicitly addressed	Yes	Self–other distinction	Real robot head	Spiking neural networks
Schillaci et al. (2013) and Schillaci (2014)	Random exploration	Not explicitly addressed	Yes	Self–other distinction	Real and simulated humanoid robot	Multi-layer perceptrons

learning sensorimotor skills. Unfortunately, links from sensorimotor development to cognitive development in these studies are often missing. Moreover, most of the abovementioned studies on intelligent exploration strategies usually address a unique sensory modality. How can exploration be performed in multimodal domains? What is the role of attention and priming behaviors in curiosity-driven exploration? Few studies tackle these issues, such as Forestier and Oudeyer (2015) and di Nocera et al. (2014). However, these and similar questions must be better addressed, in order to allow these strategies to be adopted in the implementation of more complex learning mechanisms.

In the following sections, we focus the review on studies on internal body representations and internal models, and on the predictive capabilities that they could provide to artificial agents. As it will be described in the rest of this paper, predictive processes and, in general, simulation processes of sensorimotor activity could represent the bridging mechanisms between sensorimotor learning, implemented through exploration behaviors, and the development of basic cognitive skills.
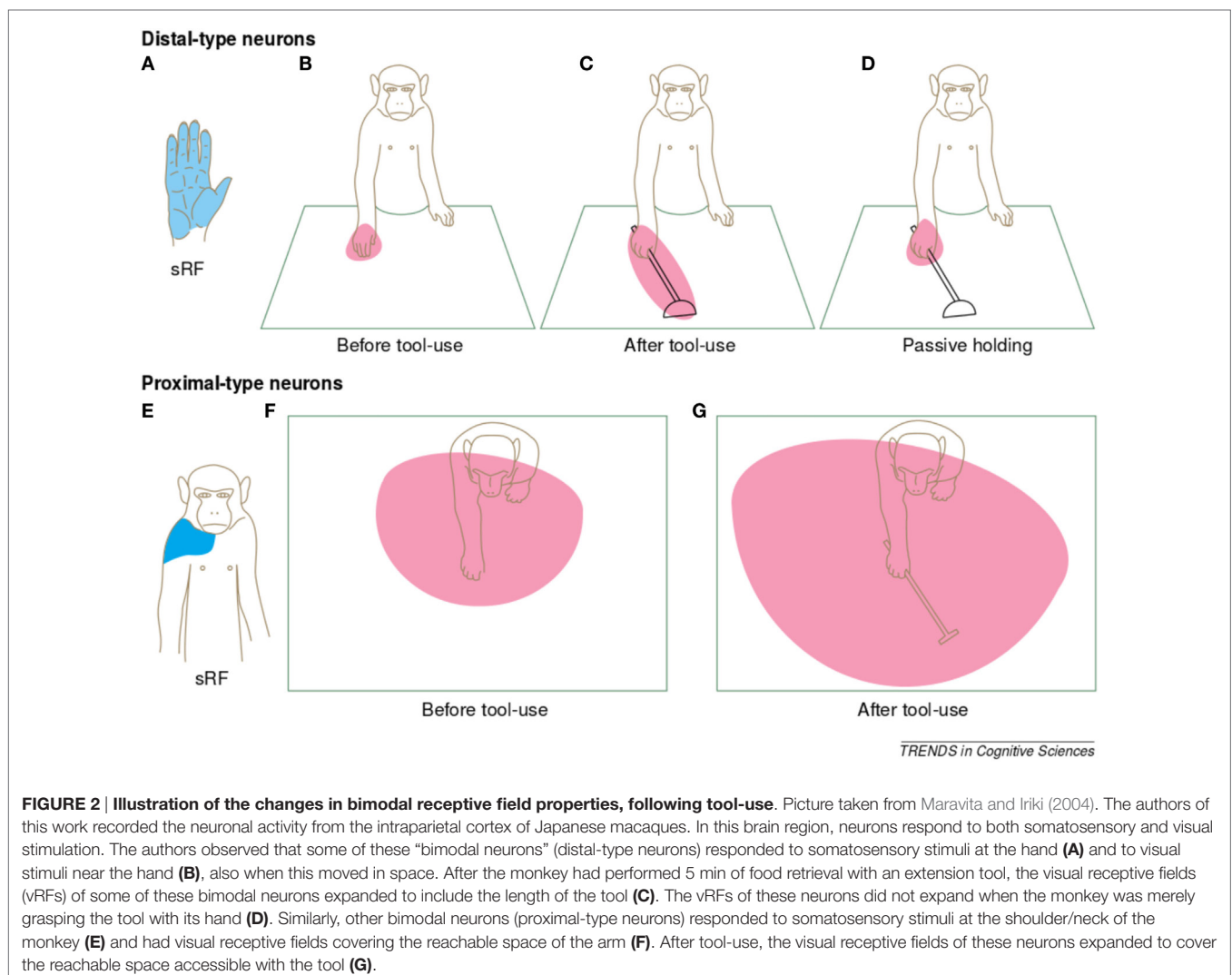
3. INTERNAL BODY REPRESENTATIONS

The rich multimodal information flowing through the sensory and motor streams during the interaction of an individual with the environment contains information about the body of the individual that has been proposed to be integrated in our brain in a sort of *body schema* (Hoffmann et al., 2010). This schema would keep an up-to-date representation of the positions of the different body parts in space and of the space of each individual modality and their combination (Hoffmann et al., 2010). Such a representation would be fundamental, for example, for constantly monitoring the position and configuration of our body, and thus for guiding our movements with respect with an environment.

In neuroscience, it is known that neural pathways and synapses in the brain change with the behavior and the interaction of the individual with the environment. Plastic changes are produced by sensory and motor experiences, which are strongly dependent on the characteristics of the body of the subject. Studies on body representations (Udin and Fawcett, 1988; Cang and Feldheim,

2013) suggested the existence of topographic maps in the brain, or projections of sensory receptors and of effector systems into structured areas of the brain. These maps self-organize throughout the brain development in a way that adjacent regions process spatially close sensory parts of the body. Kaas (1997) reported a number of studies showing the existence of such maps in the visual, auditory, olfactory, and somatosensory systems, as well as in parts of the motor brain areas. Moreover, evidence suggests that different areas belonging to different sensory and motor systems are integrated into a unique representation. The findings from Iriki et al. (1996), Maravita et al. (2003), Holmes and Spence (2004), and Maravita and Iriki (2004), for example, support the existence of an integrated representation of visual, somatosensory, and auditory peripersonal space in human and non-human primates, which operates in body-part-centered reference frames. In developmental psychology, Butterworth and Hopkins (1988) reported evidence demonstrating that various sensorimotor systems are potentially organized and coordinated in their functioning from birth, such as primitive forms of visually guided reaching (Von Hofsten, 1982). Similarly, Rochat and Morgan (1998) suggested that infants, already around the age of

12 months, possess a sense of a calibrated body schema, which is a perceptually organized entity which they can monitor and control. The existence of a body representation in the brain is also suggested by studies on sensory and motor disorders. For example, Haggard and Wolpert (2005) have shown that several sensory and motor disorders can be explained as caused by damage to some of the properties of a body representation in the human brain that are required for multimodal integration and coordinated sensorimotor control.

Body representations very likely undergo a continuous process of adaptation, as humans and animals follow an ontogenetic process, where corporal dimensions and morphology change over time. Nonetheless, even temporary alterations of the body of the individual can happen, such as those produced by the usage of tools. The way the brain deals with these changes has attracted the interest of many researchers. For example, Cardinali et al. (2009) studied the alterations in the kinematics of grasping movements from free-hand conditions to tool-use ones. Other studies (Iriki et al., 1996; Maravita and Iriki, 2004; Sposito et al., 2012; Ganesh et al., 2014) reported effects in the dynamics of movements with the usage of tools (see **Figure 2**), as well as plastic changes in

FIGURE 2 | Illustration of the changes in bimodal receptive field properties, following tool-use. Picture taken from Maravita and Iriki (2004). The authors of this work recorded the neuronal activity from the intraparietal cortex of Japanese macaques. In this brain region, neurons respond to both somatosensory and visual stimulation. The authors observed that some of these "bimodal neurons" (distal-type neurons) responded to somatosensory stimuli at the hand **(A)** and to visual stimuli near the hand **(B)**, also when this moved in space. After the monkey had performed 5 min of food retrieval with an extension tool, the visual receptive fields (vRFs) of some of these bimodal neurons expanded to include the length of the tool **(C)**. The vRFs of these neurons did not expand when the monkey was merely grasping the tool with its hand **(D)**. Similarly, other bimodal neurons (proximal-type neurons) responded to somatosensory stimuli at the shoulder/neck of the monkey **(E)** and had visual receptive fields covering the reachable space of the arm **(F)**. After tool-use, the visual receptive fields of these neurons expanded to cover the reachable space accessible with the tool **(G)**.
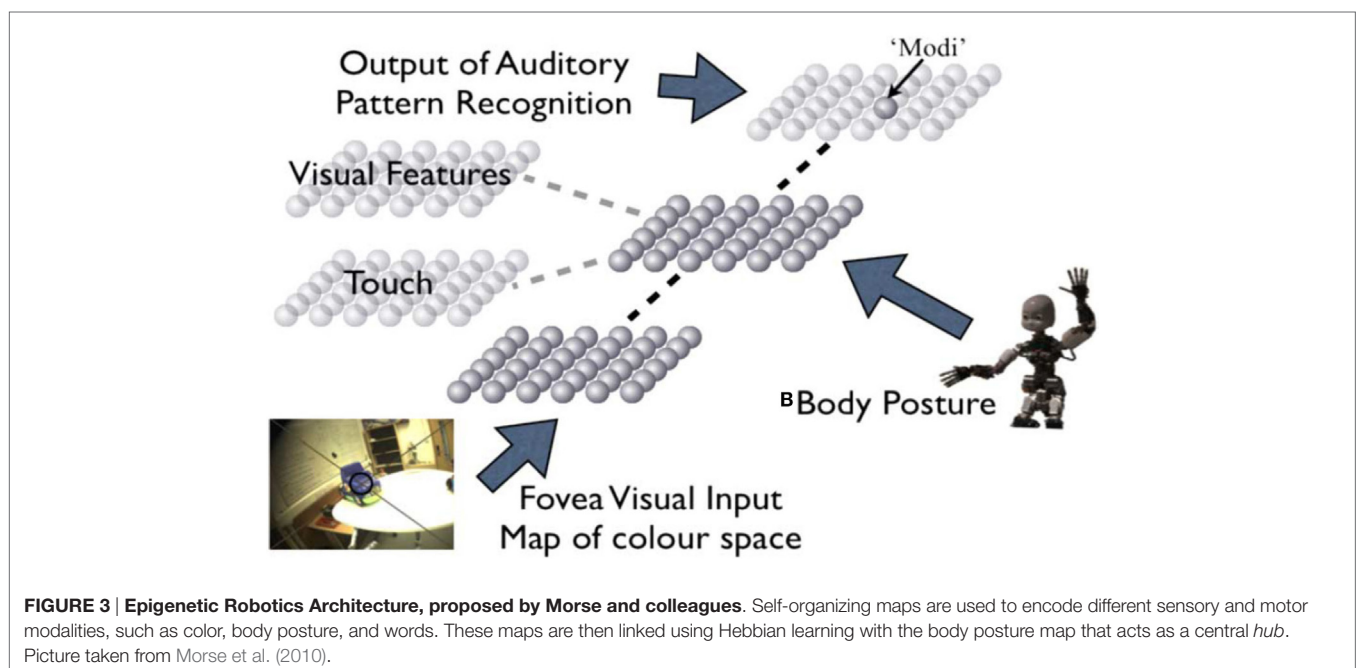
the primary somatosensory cortex in the human brain (Schaefer et al., 2004).

It is still very challenging to reproduce and to deploy in a computational model the partially unexplained but fascinating capabilities of our brain to acquire and to maintain internal body representations, and to re-adapt them to temporary or permanent bodily changes. A typical challenge is related to finding a proper balance between stability and plasticity of the internal model of the body, which can ensure both long-term memory maintenance and propensity to sudden and temporary alteration of the body schema. During the last couple of decades, interest in the possibility to develop models inspired by the mechanisms of human body representations has been growing also in the robotics community. Equipping robots with multimodal body representations, capable of adapting to dynamic circumstances, would indeed improve their level of autonomy and interactivity. Morever, body representations can be seen as the set of sensorimotor schemes that an agent acquires through the interaction with the environment.

In the robotics literature, several terms can be found referring to the same concept of the abovementioned internal body representations, such as body schemes, body maps, internal models of the body, multimodal maps, intermodal maps, and multimodal representations. The investigation in body representations in robotics has probably started within the context of the development of visuo-motor coordination. Visuo-motor coordination is often referred to as the capability to reach a particular position in the space with a robotic arm, but could also be referred to oculomotor control and eye (camera) head coordination. In both cases, this skill requires knowledge and coordination of the sensory and motor systems, thus a knowledge of an internal model or representation of the embodiment of the artificial system. For example, Metta (2000) implemented an adaptive control system inspired by human development of visuo-motor coordination for the acquisition of orienting and reaching behaviors on a humanoid robot. The robotic agent started with learning how to move its eyes only and proceeded with acquiring closed-loop gains, reflex-like modules controlling the arm sub-system, and finally eye–head and head–arm coordination. Goal-directed exploration behaviors have been compared to random exploration ones in the study. Similarly, Saegusa et al. (2008) studied the acquisition of visuo-motor coordination skills in a humanoid robot using an intelligent exploration behavior based on a prediction error-dependent interest function. Kajić et al. (2014) adopted a random exploration strategy for acquiring visuo-motor coordination skills, but proposed a biologically inspired model consisting of Self-Organizing Maps (Kohonen, 1982) for encoding the sensory and motor mapping. Such a framework led to the development of pointing gestures in the robot. The model architecture proposed by Kajić et al. (2014) was inspired by the Epigenetic Robotics Architecture [ERA, Morse et al. (2010)], where a structured association of multiple SOMs has been adopted for mapping different sensorimotor modalities in a humanoid robot. The ERA architecture resembles the formation and maintenance of topographic maps in the primate and human brain (see **Figure 3**). Shaw et al. (2015) proposed a similar architecture for body representation based on sensorimotor maps and intrinsic motivation-based exploration behaviors. In their experiment, the robot progressed through a staged development whereby eye saccades emerged first, followed by gaze control, then primitive reaching, and followed by eventual coordinated gaze-to-touch behaviors. An extension of the approach proposed by Kajić et al. (2014) was presented by Schillaci et al. (2014), where Dynamic Self-Organizing Maps [DSOMs (Rougier and Boniface, 2011)] and a Hebbian paradigm were adopted for online and continuous learning on both static and dynamic data distributions. The authors addressed the learning of visuo-motor coordination

FIGURE 3 | Epigenetic Robotics Architecture, proposed by Morse and colleagues. Self-organizing maps are used to encode different sensory and motor modalities, such as color, body posture, and words. These maps are then linked using Hebbian learning with the body posture map that acts as a central *hub*. Picture taken from Morse et al. (2010).

in robots, but focused on the capability of the proposed internal model for body representations to adapt to sudden changes in the dynamics of the system. Brandao et al. (2013) presented an architecture for integrating visually guided walking and whole-body reaching in a humanoid robot, thus increasing the reachable space that can be acquired with the visuo-motor coordination learning mechanisms proposed above. Goal-directed exploration mechanisms have been used by the authors.

Roncone et al. (2014) investigated the calibration of the parameters of a kinematic chain by exploiting the correspondences between tactile input and proprioceptive modality (joint angles), or the tactile-proprioceptive contingencies, in the humanoid robot iCub. The study is in line with the finding from Rochat and Morgan (1998), who suggested that the multimodal events continuously experienced by infants, such as the visual-proprioceptive event of looking at their own movements, or the perceptual event of the double touch resulting from the contact of two tactile surfaces, would drive the establishment of an inter-modal calibration of the body. Yoshikawa et al. (2004) addressed visuo-motor and tactile coordination in a simulated robot. In particular, they proposed a method for learning multimodal representations of the body surface through double-touching, as this co-occurred with self-occlusions. Similarly, Fuke et al. (2007) addressed the learning of a body representation consisting of motor, proprioceptive, tactile, and visual modalities in a simulated humanoid robot. The authors encoded sensory and motor modalities as self-organizing maps. Hikita et al. (2008) extended this multimodal representation to the context of tool-use in a humanoid robot. Similarly, Schillaci et al. (2012a) implemented a learning mechanisms based on random exploration strategies for the acquisition of visuo-motor coordinationon a humanoid robot (see **Figure 4**) and analyzed how the action space of both arms can vary when the robot is provided with an extension tool (see **Figures 5** and **6**). The extended arm experiment can be seen as the body of the robot being temporarily extended by a suitable tool for a specific task (Schillaci, 2014).

Nonetheless, several other studies can be found in the literature, which address body representations for artificial agents outside the context of visuo-motor coordination. For example, Hafner and Kaplan (2008) extended the notion of body representations, or body maps, to that of *interpersonal maps*, a geometrical representation of the relationships between a set of proprioceptive and heteroceptive information sources. The study proposed a common representation space for comparing an agent's behavior and the behavior of other agents, which was used to detect specific types of interactions between agents, such as imitation, and to implement a prerequisite for affordance learning. The abovementioned Epigenetic Robotics Architecture Morse et al. (2010) addressed body representations for grounding linguistic labels onto body postures, visual, and auditory modalities. A similar framework has been proposed by Lallee and Dominey (2013), which encodes sensory and motor modalities as self-organizing maps into a body representation. Through the use of a goal-directed exploration behavior, the system learns a body model composed of specific modalities (arm proprioception, gaze proprioception, vision) and their multimodal mappings, or contingencies. Once multimodal mappings have been learned,
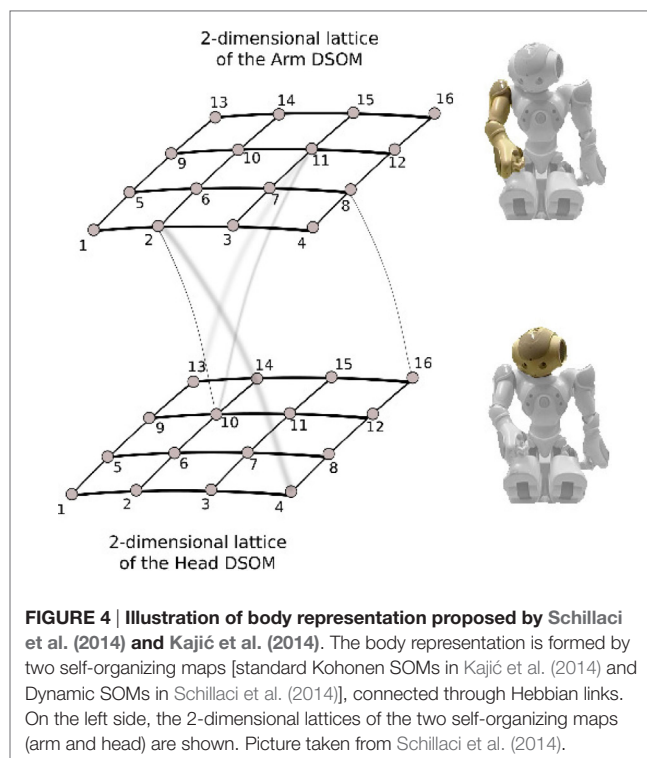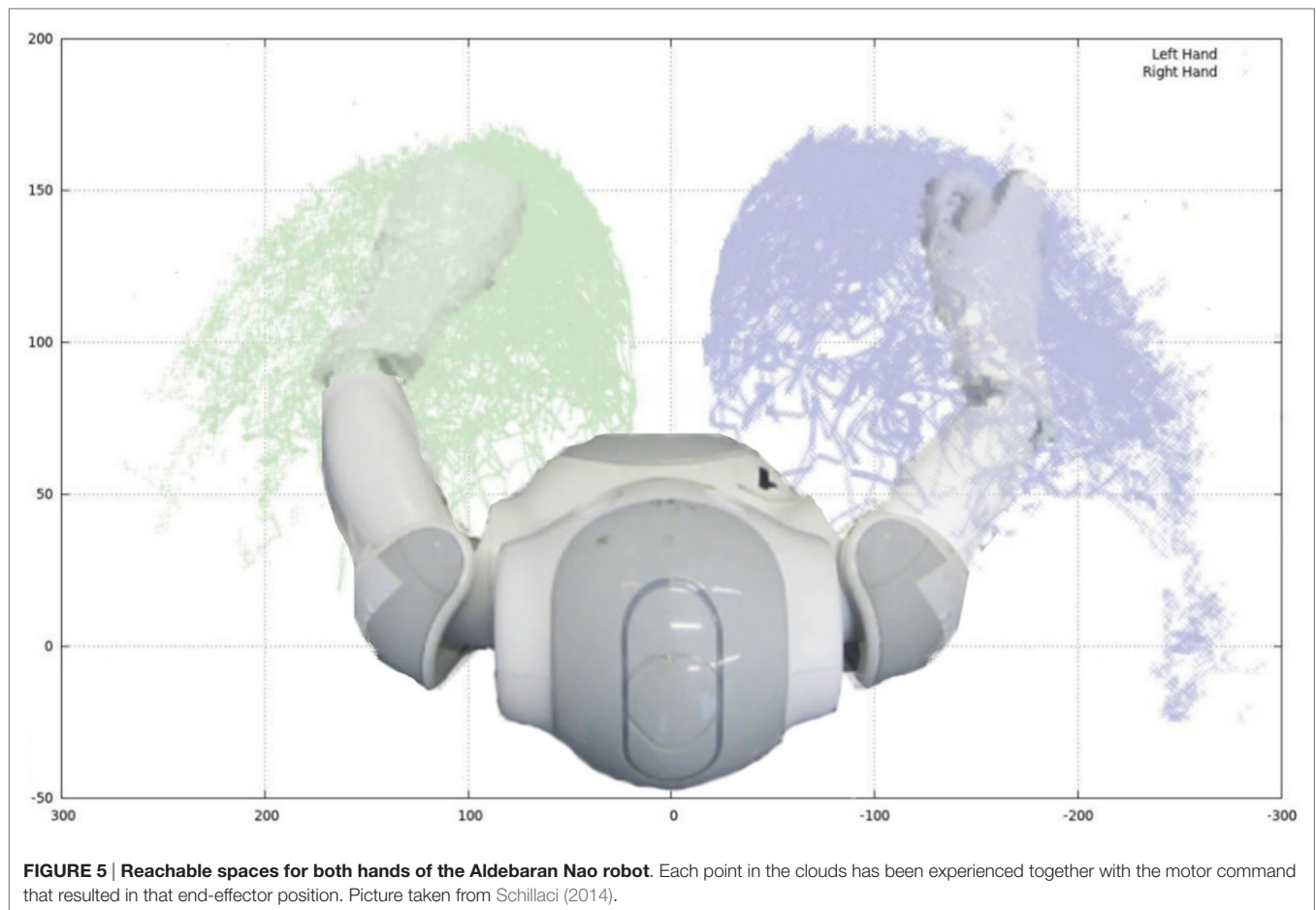

FIGURE 4 | Illustration of body representation proposed by Schillaci et al. (2014) and Kajić et al. (2014). The body representation is formed by two self-organizing maps [standard Kohonen SOMs in Kajić et al. (2014) and Dynamic SOMs in Schillaci et al. (2014)], connected through Hebbian links. On the left side, the 2-dimensional lattices of the two self-organizing maps (arm and head) are shown. Picture taken from Schillaci et al. (2014).

the system is capable of generating and exploiting internal representations or mental images based on inputs in one of these multiple dimension (Lallee and Dominey, 2013). Kuniyoshi and Sangawa (2006) presented a model of neuro-musculo-skeletal system of a human infant, composed of self-organizing cortical areas for primary somatosensory and motor areas that participate in the explorative learning by simultaneously learning and controlling the movement patterns. In the simulated experiment, motor behaviors emerged, including rolling over and crawling-like motion. Body representations that include the auditory modality have been also addressed, although not explicitly, by Ince et al. (2009), who investigated methods for the prediction and suppression of ego-motion noise. The authors built up an internal body representation of a humanoid robot consisting of motor sequences mapped to the recorded motor noise and their spectra. This resulted in a large noise template database that was then used for ego-noise prediction and subtraction.

Exploration behaviors for the acquisition and maintenance of internal body representations is a very elegant and promising developmental approach for providing artificial agents with robustness and adaptivity to dynamic body and environments. However, how can these low-level behaviors and representations enable the development of more complex cognitive and motor capabilities? Although this question has still not been clearly answered, several behavioral and brain studies suggest that processes of mental simulations of action–perception loops are likely to be executed in our brain and are dependent on internal motor representations. The capability to simulate sensorimotor experience might represent a key mechanism behind the implementation of higher cognitive skills, as discussed in the following section.

FIGURE 5 | Reachable spaces for both hands of the Aldebaran Nao robot. Each point in the clouds has been experienced together with the motor command that resulted in that end-effector position. Picture taken from Schillaci (2014).
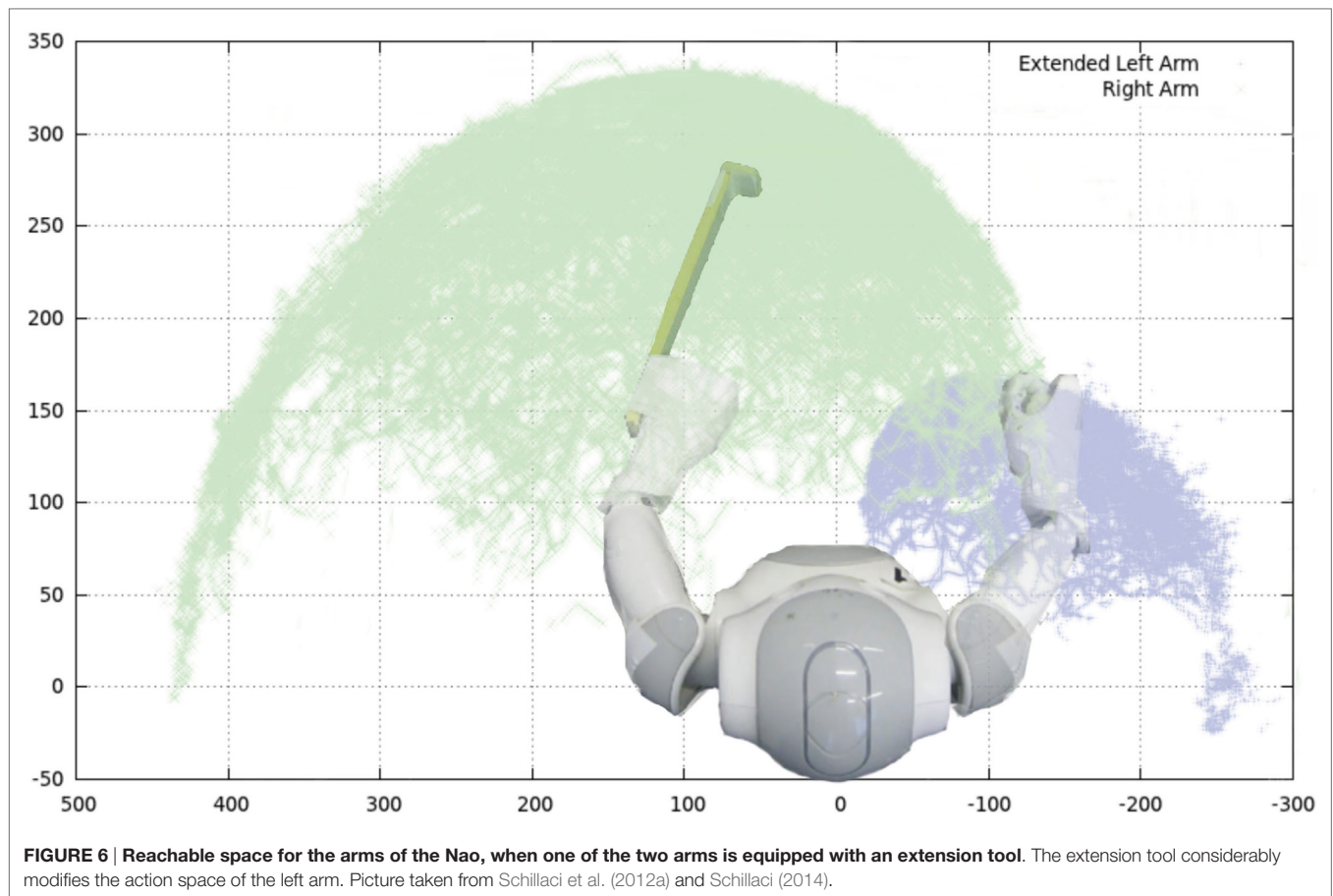
4. SENSORIMOTOR SIMULATIONS

In one of the most influential post-cognitivist studies, Lakoff and Johnson (1980) argued that cognitive processes are expressed and influenced by metaphors, which are based on personal experiences and shape our perceptions and actions. Correlations and co-occurrence of embodied experiences would lead to primitive conceptual metaphors. As argued by Lakoff, physical concepts, such as running and jumping, can be understood through the sensorimotor system, as they can be performed, seen, and felt. Abstract concepts would get their meaning via conceptual metaphors, a combination of basic primitive metaphors that get their meaning via embodied experience. Therefore, Lakoff (2014) concludes that the meaning of concepts comes through embodied cognition. Moreover, in Lakoff and Johnson (1980), the authors argued that metaphorical inferences would arise from neural simulation of experienced situations.

Similarly, Varela, Thompson, and Rosch argued that the interactions between the body, its sensorimotor circuit, and the environment determine the way the world is experienced. Cognitive agents are living bodies *situated* in the environment and knowledge would emerge through the *embodied* interaction with the world (Varela et al., 1992). According to the *enaction* paradigm proposed by Varela and colleagues, the embodied actions of an individual in the world constitute the way how the environment is experienced and thereby ground the agent's cognition. This is at least accepted in the Narrow Conception of Enactivism (de Bruin and Kästner, 2012).

A related concept is known in the philosophical and scientific literature as *mental imagery* [for a literature review on embodied cognition and mental imagery, see Schillaci (2014)]. This phenomenon has been defined as a quasi-perceptual experience (in any sensory modality, such as auditory, olfactory, and so on) which resembles perceptual experience but occurs in absence of external stimuli (Nigel, 2014). What is the nature of this mental phenomenon has always been a very debated topic [Nigel (2014) provides a more comprehensive review of the literature on mental imagery]. Not surprisingly, studies on mental imagery can be found already in Greek philosophy. In *De Anima*, Aristotle saw mental images, residues of actual impressions or *phantasmata* as playing a central role in human cognition, for example, in memory. Behaviorists believed that psychology must have handled only observable behaviors of people and animals, not unobservable introspective events. Therefore, mental imagery was reputed as not being sufficiently scientific (Watson, 1913), since no rigorous experimental method was proposed to demonstrate it. Only after the 1960s, mental imagery gained new attention in psychology and in the neuroscience (Nigel, 2014).

FIGURE 6 | Reachable space for the arms of the Nao, when one of the two arms is equipped with an extension tool. The extension tool considerably modifies the action space of the left arm. Picture taken from Schillaci et al. (2012a) and Schillaci (2014).

During the last 20 years, many behavioral and cognitive studies on attitudes, emotion, and social perception investigated and supported the hypothesis that the body is closely tied with cognition. We argue that sensorimotor simulations are behind all of these processes. Strack et al. (1988) demonstrated that people's facial activity influences their affective responses. Participants were holding a pen in their mouth in a way that either inhibited or facilitated the muscles typically associated with smiling without requiring subjects to pose in a smiling face. The authors found that subjects reported more intense humor responses when cartoons were presented under facilitating conditions than under inhibiting conditions (Strack et al., 1988). These results highlight the important overlapping between motor activity and the affective response an agent has.

Wexler and Klam (2001) presented a study where participants predicted the position of moving objects, in cases of actively produced and passively observed movement. The authors found that in the absence of eye tracking, when occluding the object, the estimates are more anticipatory in the active conditions than in the passive ones. The anticipatory effect of an action depended on the congruence between the motor action and the visual feedback: the less congruent were the motor action and the visual feedback, the more diminished the anticipatory effect, but it was never eliminated. However, when the target was only visually tracked, the effect of manual action

disappeared, indicating distinct contributions of hand and eye movement signals to the prediction of trajectories of moving objects (Wexler and Klam, 2001).

Animal research also suggests that rat brains implement simulation processes. O'Keefe and Recce (1993) found that particular cells in the hippocampus of the rat's brain seem to be involved in the representation of the animal's position. Their observations of the firing characteristics of these cells suggested that the position of the animal is periodically anticipated along the path. In a study on visual guidance of movements in primates (Eskandar and Assad, 1999), monkeys were trained to use a joystick to move a spot to a specific target. During the movements, the authors modified the relationship between the direction of joystick and movements of the spot, and eventually occluded the spot, thus dissociating the visual and motor correlations. The authors observed cells in the lateral intraparietal area of the monkey's brain, which were not selectively modulated by either visual input or motor output, but rather seemed to encode the predicted visual trajectory of the occluded target (Eskandar and Assad, 1999).

Wolpert et al. (1995) suggested that sensorimotor prediction processes exist in motor planning and execution also in humans. In testing whether the central nervous system is able to maintain an estimate of the position of the limbs, the authors asked participants to move their arm in the absence of visual feedback. Each participant gripped a tool that was used to measure the position

of the thumb and to apply forces to the hand using torque motors. The experimenters were disturbing the hand movements of the participants, which were then asked to indicate the visual estimate of the unseen thumb position using a trackball held, in the other hand. The distance between the actual and visual estimate of thumb location, used as a measure of the state estimation error, showed a consistent overestimation of the distance moved (Wolpert et al., 1995). The authors observed a systematic increase of the error during the first second of movement and then a decay. Therefore, they proposed that the initial phase is the result of a predictive process that estimates the hand position, followed by a correction of the estimate when the proprioceptive feedback is available (Wolpert et al., 1995). In another study, Wolpert et al. (1998) suggested that an internal body representation consisting of a combination of sensory input and motor output signals is stored in the posterior parietal cortex of the brain. The authors also reported that a patient with a lesion of the superior parietal lobe showed both sensory and motor deficits consistent with an inability to maintain such an internal representation between updates.

Blakemore et al. (2000b) supported the existence of self-monitoring mechanisms in the human brain for explaining why tickling sensations cannot be self-produced. The proposal is that sensory consequences of self-generated actions are perceived differently from an identical sensory input that is externally generated. This would explain the cancelation or attenuated tickle sensation when this is the consequence of self-produced motor commands (Blakemore et al., 2000b). The data reported in the study suggest that brain activity differs in response to externally and internally produced stimuli. Moreover, it has been proposed that illnesses, such as schizophrenia, would disable the patient's capability to detect self-produced actions, therefore producing an altered perception of the world (Frith et al., 2000).

The internal models proposed by Wolpert et al. (1998) could explain the computational processes behind the attenuation of sensory sensation reported above. In particular, Wolpert and colleagues suggest that these internal models are constructed through the sensorimotor experience of the agent in the environment and used in simulation for processes, such as the attenuation of sensory sensations in Blakemore et al. (2000a) and conditions as in Frith et al. (2000). A similar effect has been reported by Weiss and colleagues in a study on selective attenuation of self-generated sounds (Weiss et al., 2011). The experience of generating actions, or self-agency, has been suggested to be linked to the internal motor signals associated with the ongoing actions. It has been proposed that the experience of perceiving actions as self-generated would be caused by the anticipation and, thus, the attenuation of the sensory consequences of such motor commands (Weiss et al., 2011). The results reported by the authors confirmed this hypothesis, as they found that participants perceived the loudness of sounds less intensive when they were self-generated than when they were generated by another person or by a software.

Further evidence suggesting that an internal model of our motor system is involved in the capability to distinguish between self and others can be found in Casile and Giese (2006). In this study, the authors showed that participants were better at
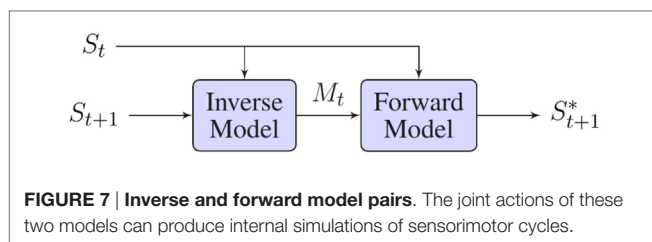
recognizing themselves than others when watching movies of only point-light walkers. Knoblich and Flach (2001) performed a study on the capability of participants to predict the landing position of a thrown dart, observed from a video screen. The authors reported that predictions were more accurate when participants observed their own throwing actions than when they observed another person's throwing actions, even if the stimulus displays were exactly the same for all participants. The results are consistent with the assumption that perceptual input can be linked with the action system to predict future outcomes of actions (Knoblich and Flach, 2001).

4.1. Computational Models for Sensorimotor Simulations

Hesslow (2002) supported with a set of evidence the *simulation theory of conscious thought*, by assuming that simulation processes are implemented in our brain and that the simulation approach can explain the relations between motor, sensory, and cognitive functions and the appearance of an inner world. In the investigation on internal simulation processes in the human brain, internal forward and inverse models have been proposed (Wolpert et al., 2001). A forward model is an internal model which incorporates knowledge about sensory changes produced by self-generated actions of an individual. In other words, a forward model predicts a sensory outcome S_{t+1} of a motor command M_t applied from an initial sensory situation S_t. This internal model was first proposed in the control literature as a means to overcome problems, such as the delay of feedback on standard control strategies and the presence of noise, both also characteristic of natural systems (Jordan and Rumelhart, 1992). More recently, Webb (2004) presented a discussion on the possibilities offered by the studies in invertebrate neuroscience to unveil the existence of these types of models. The research concludes that although there is no conclusive evidence, forward models might answer some of the open questions on the mapping between motor and sensory information.

While forward models present the causal relation between actions and their consequences, inverse models perform the opposite transformation providing a system with the necessary motor command M_t to go from a current sensory situation S_t to a desired one (S_{t+1}) (see **Figure 7**). Inverse models are also very well known in control theory and in robotics, as they have been used for the implementation of inverse kinematics in robotic manipulators. Kinematics describe the geometry of motion of points and objects. In classic control theory, kinematics equations are used to determine the joint configuration of a robot to reach a desired position of its end-effector.

Recently, forward and inverse models became central players in the coding of sensorimotor simulations, as they naturally fuse together different sensory modalities as well as motor information, not only providing individuals with multimodal representations but also encoding the dynamics of their motor systems (Wilson and Knoblich, 2005). Studies such as the ones reported in the previous section shed light on the importance that the prediction of the sensory consequences of our own actions has for basic motor tasks (Blakemore et al., 1998). Forward

FIGURE 7 | Inverse and forward model pairs. The joint actions of these two models can produce internal simulations of sensorimotor cycles.

models, by functioning with self-generated motor commands are an important base for the feeling of agency, as suggested by Weiss et al. (2011). A faulty functioning of forward models, in their role as self-monitoring mechanisms, is thought to be responsible for some of the symptoms present in schizophrenia (Frith, 1992). In general, the capability to *anticipate* sensorimotor activity is thought to be crucially involved in several cognitive functions, including attention, motor control, planning, and goal-oriented behavior (Pezzulo, 2007; Pezzulo et al., 2011).

Research has been done on computational internal models for action preparation and movement, in the context of reaching objects and of handling objects with different weights (Wolpert and Ghahramani, 2000). The main proposal became a standard reference known as the MOdular Selection And Identification for Control (MOSAIC) model (Haruno et al., 2001). In MOSAIC, different pairs of inverse and forward models encode specific sensorimotor schemes. The contribution of each pair to choose a motor command is weighted by a responsibility estimator according to the context and the behavior the system is currently modeling (Haruno et al., 2001). The authors extended the model to encode more complex behaviors and actions in Hierarchical MOSAIC (Wolpert et al., 2003). Conceptually, HMOSAIC is capable of accounting and model social interaction, action observation, and action recognition.

Tani et al. (2005) proposed an architecture in which multiple sensorimotor schemes can be learned in a distributed manner based on using a recurrent neural network with parametric biases. The model was demonstrated to implement behavior generation and recognition processes in an imitative interaction experiment, thus acting as a mirror system. Moreover, the model has been shown to support associative learning between behaviors and language, supporting the hypothesis posed by Arbib (2002) that the capabilities of the mirror neurons for conceptualizing objects manipulation behaviors might lead to the origins of language (Tani et al., 2005). In the framework of cognitive robotics, interesting work has been done in incorporating internal simulations for navigation on autonomous robots. Ziemke et al. (2005) incorporated several aspects of the sensorimotor theories and performed internal simulations to achieve a navigation task. A trained robot equipped with the proposed framework was able, in some cases, to move blindly in a simple environment, using as input only own sensory predictions rather than actual sensory input.

Lara and Rendon-Mancha (2006) equipped a simulated agent with a forward model implemented as an artificial neural network. The system learned to successfully predict multimodal sensory representations formed by visual and tactile stimuli

for an obstacle avoidance task. Following the same strategy, Escobar et al. (2012) made an experiment on robot navigation through self body-mapping and the association between motor commands and their respective sensory consequences. A mobile robot was made to interact with its environment in order to know the free space around it from re-enaction of sensory–motor cycles predicting collisions from visual data. The robot formed multimodal associations, consisting of motor commands, disparity maps from vision and tactile feedback, into a forward model, which was trained with data coming from random trajectories. The resulting forward model allowed the robot to navigate avoiding undesired situations by performing long-term predictions of the sensory consequences of its actions (Escobar et al., 2012).

Following navigation studies, Möller and Schenck (2008) made an experiment on anticipatory dead-end recognition, where a simulated agent learned to distinguish between dead ends and corridors without the necessity to represent these concepts in the sensory domain. With interacting with the environment, the agent acquired a visuo-tactile forward model that allowed it to predict how the visual input was changing under its movements and whether movements were leading to a collision. In addition, the agent learned an inverse model for suggesting which actions should be simulated for long-term predictions. Finally, Hoffmann and Möller (2004) and Hoffmann (2007) presented a chain of forward models that provides a mobile agent with the capability to select different actions to achieve a goal situation and perform mental transformations during navigation. It is worth highlighting that in the last five examples, the agents make use of long-term predictions (LTP) to achive the desired behaviors. These LTPs are achieved by executing sensorimotor simulations aquired throught the interaction of the agents with the environment.

Akgün et al. (2010) presented an internal simulation mechanism for action recognition, inspired by the behavior recognition hypothesis of mirror neurons. The proposed computational model, similar to HAMMER and MOSAIC, is capable of recognizing actions online using a modified Dynamical Movement Primitives framework, a non-linear dynamic system that has been proposed for imitation learning, action generation, and recognition by Ijspeert et al. (2001). Schrodt et al. (2015) presented a generative neural network model for encoding biological motion, for recognizing observed movements and for adopting the point of view of an observer. The proposed model learns map and segment multimodal sensory streams of self-motion to anticipate motion progression, to complete missing sensory information, and to self-generate motion sequences that have been previously learned. In addition, the model was equipped with the capability to adopt the point of view of an observed person, establishing full consistency with the embodied self-motion encodings by means of active inference (Schrodt et al., 2015).

A MOSAIC-like architecture for action recognition was also presented by Schillaci et al. (2012b), where the authors also compared different learning strategies for inverse and forward model pairs (see **Figure 8**). In an experiment on action selection, Schillaci et al. (2012b, 2014) showed how a robot can deal with tool-use when equipped with self-exploration behaviors and
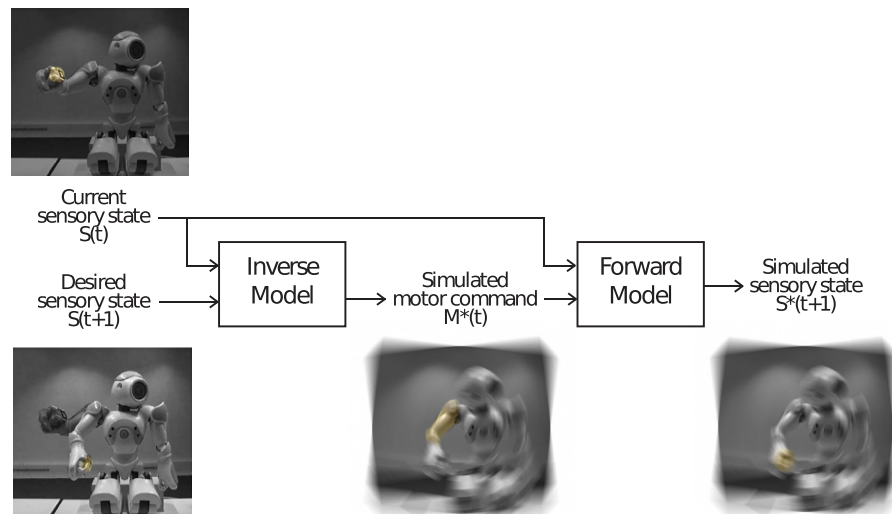
FIGURE 8 | An example of an internal simulation [image taken from Schillaci (2014)]. The inverse model simulates the motor command (in the example, a displacement of the joints of one arm of the humanoid robot Aldebaran Nao) needed for reaching a desired sensory state, from the current state of the system. Before being sent to the actuators, such a simulated motor command can be fed into the forward model that anticipates its outcome, in terms of sensory perception. A prediction error of the internal simulation can be calculated by comparing the simulated sensory outcome with the desired sensory state.

with the capability to execute internal simulations of sensorimotor cycles. Schillaci et al. implemented learning of internal models through self-exploration on a humanoid robot, which were consequently used for predicting simple arm trajectories and for distinguishing between self-generated movements and arm trajectories executed by a different robot (Schillaci et al., 2013) or by a human (Schillaci, 2014).

Interesting research on sensorimotor simulations can be found in the context of action execution and recognition. For example, Dearden and Demiris (2005) presented a study where a robot learned a forward model that successfully imitated actions presented to its visual system. In a later study, Dearden (2008) presented a more complex system where a robot learns from a social context by means of forward and inverse models using memory-based approaches. Nishide et al. (2007) presented a study on predicting object dynamics through active sensing experiences with a humanoid robot. For predicting the movements of an unknown object, a static image of the object and robot motor command are fed into a neural network that was trained in a previous stage through a learning mechanism based on active sensing. In the HAMMER architecture (Hierarchical Attentive Multiple Models for Execution and Recognition) proposed by Demiris and Khadhouri (2006), inverse and forward model pairs encoded sensorimotor schemes and were used for action execution and action understanding. The HAMMER architecture was implemented using Bayesian Belief Networks and was also extended to include cognitive processes, such as attention (Demiris and Khadhouri, 2006).

Kaiser (2014) investigated a computational model for perceiving the functional role of objects, or their affordances, based on internally simulated object interactions. The approach was based on an implementation of visuo-motor forward

models based on feed-forward neural networks and geometric approximations. The models were trained with sensorimotor data gathered from self-exploration, although in a structured systematic fashion, i.e., by defining grids in sensorimotor space or in motor space (Kaiser, 2014).

A promising line of investigation addresses the implementation of simulation processes for the development of the sense of agency, the sense of being the cause or author of a movement, and for distinguishing between self and other. Pitti et al. (2009) proposed a mechanism of spike timing-dependent synaptic plasticity as a biologically plausible model for detecting contingency between multimodal events and for allowing a robotic agent to experience its own agency during motion.

Finally, we would like to highlight the work presented in Hoffmann (2014), where the paradigm of cognitive developmental robotics is addressed through a case study. In this, information flow is analyzed with an agent interacting with the world. A very critical view of the paradigm is addressed in the light of embodied cognition and the enactive paradigm. Extraction of low-level features in the sensorimotor space is analyzed and use in higher level behaviors of the agent where sensorimotor associations are formed. Interestingly, an important conclusion is the importance and usefullness of forward models in the control structure of agents.

5. CONCLUSION

The goal of developmental roboticists is to implement mechanisms for autonomous motor and mental development in artificial agents. We argued that mechanisms for sensorimotor simulation may be the bridge between low-level sensorimotor representations learned through experience and the implementation of

basic cognitive skills in artificial agents. Several robotics studies showed that internal simulations and imagery can provide robots with capabilities, such as long-term prediction for navigation, behavior selection and recognition, and perception of the functional role of objects, and can even serve as a possible basis for the acquisition of the sense of agency and for the capability to distinguish between self and other.

A prerequisite for the implementation of sensorimotor simulation processes in artificial agents is the knowledge about the characteristics of their motor systems and their embodiment. In fact, to be able to internally simulate the outcome of their own actions, robots need to know their action possibilities and to have an antecedent perceptual experience about the consequences of their activities. An elegant and promising way for allotting artificial agents with such a knowledge is provided by exploration, a learning mechanism inspired by human development. By exploring their bodily capabilities and by interacting with the environment, possibly using mechanisms resembling human curiosity, robots can generate a rich amount of sensory and motor experience. Maintaining this multimodal information into internal representations of the robot's body could be not only helpful for monitoring the correct functioning of the system but also exploited for detecting unexpected events, such as temporary or permanent changes in the agents morphology, and for adapting to them. Such a possibility would be impossible to implement with *a priori* defined models of the robot body and its surrounding environment, as this would require not only the exact knowledge of the dynamics of the artificial system and its surroundings, as well as the definition of all the variables that could affect the normal functioning of the system. It is important to note that different implementations have made use of different computational strategies for the coding of these body representations. However, in all cases, these representations encompass the bulk of the possibilities an agent has of sensing and acting in the world. Following this line of thought, simulations are the off-line rehearsal of these schemes.

We argue that sensorimotor learning, internal body representation, and internal sensorimotor simulations are paramount in the development of artificial agents. Also, we strongly believe that the three processes have to be considered interdependent and necessary when investigating autonomous mental development. It is for these reasons, we tried to give an interdisciplinary overview of what we believe to be the most prominent studies on these topics, from the disciplines of robotics, cognitive sciences, and neuroscience.

AUTHOR CONTRIBUTIONS

Each of the authors has contributed equally and significantly to the study.

ACKNOWLEDGMENTS

REFERENCES

Akgün, B., Tunaoglu, D., and Sahin, E. (2010). "Action recognition through an action generation mechanism," in *International Conference on Epigenetic Robotics (EPIROB)*. Örenäs Slott.

Arbib, M. A. (2002). "The mirror system, imitation, and the evolution of language," in *Imitation in Animals and Artifacts*, eds K. Dautenhahn and C. L. Nehaniv (Cambridge, MA: MIT Press), 229–280.

Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., et al. (2009). Cognitive developmental robotics: a survey. *IEEE Trans. Auton. Ment. Dev.* 1, 12–34. doi:10.1109/TAMD.2009.2021702

Baranes, A., and Oudeyer, P.-Y. (2013). Active learning of inverse models with intrinsically motivated goal exploration in robots. *Rob. Auton. Syst.* 61, 49–73. doi:10.1016/j.robot.2012.05.008

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi:10.1146/annurev.psych.59.103006.093639

Berlyne, D. (1960). *Conflict, Arousal and Curiosity*. New York, NY: McGraw-Hill.

Berthold, O., and Hafner, V. V. (2015). "Closed-loop acquisition of behaviour on the sphero robot," in *Proceedings of the European Conference on Artificial Life* (York, UK).

Blakemore, S. J., Goodbody, S. J., and Wolpert, D. M. (1998). Predicting the consequences of our own actions: the role of sensorimotor context estimation. *J. Neurosci.* 18, 7511–7518.

Blakemore, S.-J., Smith, J., Steek, R., Johnstone, E. C., and Frith, C. D. (2000a). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring. *Psychol. Med.* 30, 1131–1139. doi:10.1017/S0033291799002676

Blakemore, S. J., Wolpert, D., and Frith, C. (2000b). Why can't you tickle yourself? *Neuroreport* 11, 11–16. doi:10.1097/00001756-200008030-00002

Brandao, M., Jamone, L., Kryczka, P., Endo, N., Hashimoto, K., and Takanishi, A. (2013). "Reaching for the unreachable: integration of locomotion and whole-body movements for extended visually guided reaching," in *2013 13th IEEE-RAS International Conference on Humanoid Robots (Humanoids)* (Atlanta: IEEE), 28–33.

Butterworth, G., and Hopkins, B. (1988). Hand-mouth coordination in the new-born baby. *Br. J. Dev. Psychol.* 6, 303–314. doi:10.1111/j.2044-835X.1988.tb01103.x

Cang, J., and Feldheim, D. A. (2013). Developmental mechanisms of topographic map formation and alignment. *Annu. Rev. Neurosci.* 36, 51–77. doi:10.1146/annurev-neuro-062012-170341

Cardinali, L., Frassinetti, F., Brozzoli, C., Urquizar, C., Roy, A. C., and Farn, A. (2009). Correspondence. *Curr. Biol.* 19, R478–R479. doi:10.1016/j.cub.2009.05.009

Casile, A., and Giese, M. A. (2006). Nonvisual motor training influences biological motion perception. *Curr. Biol.* 16, 69–74. doi:10.1016/j.cub.2005.10.071

de Bruin, L. C., and Kästner, L. (2012). Dynamic embodied cognition. *Phenomenol. Cogn. Sci.* 11, 541–563. doi:10.1007/s11097-011-9223-1

Dearden, A. (2008). *Developmental Learning of Internal Models for Robotics*. Ph.D. thesis, Imperial College London, London.

Dearden, A., and Demiris, Y. (2005). "Learning forward models for robots," in *Int. Joint Conferences on Artificial Intelligence* (Edinburgh: IEEE), 1440.

Demiris, Y., and Khadhouri, B. (2006). Hierarchical attentive multiple models for execution and recognition of actions. *Rob. Auton. Syst.* 54, 361–369. doi:10.1016/j.robot.2006.02.003

di Nocera, D., Finzi, A., Rossi, S., and Staffa, M. (2014). The role of intrinsic motivations in attention allocation and shifting. *Front. Psychol.* 5:273. doi:10.3389/fpsyg.2014.00273

Escobar, E., Hermosillo, J., and Lara, B. (2012). "Self body mapping in mobile robots using vision and forward models," in *2012 IEEE Ninth Electronics, Robotics and Automotive Mechanics Conference (CERMA)* (Cuernavaca: IEEE), 72–77.

Eskandar, E. N., and Assad, J. A. (1999). Dissociation of visual, motor and predictive signals in parietal cortex during visual guidance. *Nat. Neurosci.* 2, 88–93. doi:10.1038/4594

Forestier, S., and Oudeyer, P.-Y. (2015). "Towards hierarchical curiosity-driven exploration of sensorimotor models," in *2015 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)* (Providence: IEEE), 234–235.

Frith, C. D. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Hove, UK; Hillsdale, USA: L. Erlbaum Associates.

Frith, C. D., Blakemore, S.-J., and Wolpert, D. M. (2000). Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Res. Rev.* 31, 357–363. doi:10.1016/S0165-0173(99)00052-1

Fuke, S., Ogino, M., and Asada, M. (2007). Body image constructed from motor and tactile images with visual information. *Int. J. HR* 4, 347–364. doi:10.1142/S0219843607001096

Ganesh, G., Yoshioka, T., Osu, R., and Ikegami, T. (2014). Immediate tool incorporation processes determine human motor planning with tools. *Nat. Commun.* 5, 4524. doi:10.1038/ncomms5524

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., and Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends Cogn. Sci.* 17, 585–593. doi:10.1016/j.tics.2013.09.001

Hafner, V. V., and Kaplan, F. (2008). "Interpersonal maps: how to map affordances for interaction behaviour," in *Towards Affordance-Based Robot Control, Volume 4760 of Lecture Notes in Computer Science*, eds E. Rome, J. Hertzberg, and G. Dorffner (Berlin, Heidelberg: Springer), 1–15.

Hafner, V. V., and Schillaci, G. (2011). From field of view to field of reach – could pointing emerge from the development of grasping? *Front. Comput. Neurosci.* 5, 17 doi:10.3389/conf.fncom.2011.52.00017

Haggard, P., and Wolpert, D. M. (2005). "Disorders of body scheme," in *Higher-Order Motor Disorders*, eds H. J. Freund, M. Jeannerod, M. Hallett, and R. Leiguarda (Oxford: University Press).

Haruno, M., Wolpert, D. M., and Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Comput.* 13, 2201–2220. doi:10.1162/089976601750541778

Hesslow, G. (2002). Conscious thought as simulation of behaviour and perception. *Trends Cogn. Sci.* 6, 242–247. doi:10.1016/S1364-6613(02)01913-7

Hikita, M., Fuke, S., Ogino, M., Minato, T., and Asada, M. (2008). "Visual attention by saliency leads cross-modal body representation," in *7th IEEE International Conference on Development and Learning, 2008. ICDL 2008* (Monterey, CA: IEEE), 157–162.

Hoffmann, H. (2007). Perception through visuomotor anticipation in a mobile robot. *Neural Networks* 20, 22–33. doi:10.1016/j.neunet.2006.07.003

Hoffmann, H., and Möller, R. (2004). "Action selection and mental transformation based on a chain of forward models," in *Proc. of the 8th Int. Conference on the Simulation of Adaptive Behavior* (Cambridge, MA: MIT Press), 213–222.

Hoffmann, M. (2014). "Minimally cognitive robotics: body schema, forward models, and sensorimotor contingencies in a quadruped machine," in *Contemporary Sensorimotor Theory* (Berlin: Springer), 209–233.

Hoffmann, M., Marques, H., Hernandez Arieta, A., Sumioka, H., Lungarella, M., and Pfeifer, R. (2010). Body schema in robotics: a review. *IEEE Trans. Auton. Ment. Dev.* 2, 304–324. doi:10.1109/TAMD.2010.2086454

Holmes, N., and Spence, C. (2004). The body schema and multisensory representation(s) of peripersonal space. *Cogn. Process.* 5, 94–105. doi:10.1007/s10339-004-0013-3

Ijspeert, A., Nakanishi, J., and Schaal, S. (2001). "Trajectory formation for imitation with nonlinear dynamical systems," in *Proceedings. 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2001*, Vol. 2 (Outrigger Wailea Resort: IEEE), 752–757.

Ince, G., Nakadai, K., Rodemann, T., Hasegawa, Y., Tsujino, H., and Imura, J. (2009). "Ego noise suppression of a robot using template subtraction," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2009. IROS 2009* (St. Louis, MO: IEEE), 199–204.

Iriki, A., Tanaka, M., and Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque postcentral neurones. *Neuroreport* 7, 2325–2330. doi:10.1097/00001756-199610020-00010

Jauffret, A., Cuperlier, N., Gaussier, P., and Tarroux, P. (2013). From self-assessment to frustration, a small step towards autonomy in robotic navigation. *Front. Neurorobot.* 7, 16. doi:10.3389/fnbot.2013.00016

Jordan, M. I., and Rumelhart, D. E. (1992). Forward models: supervised learning with a distal teacher. *Cogn. Sci.* 16, 307–354. doi:10.1207/s15516709cog1603_1

Kaas, J. H. (1997). Topographic maps are fundamental to sensory processing. *Brain Res. Bull.* 44, 107–112. doi:10.1016/S0361-9230(97)00094-4

Kaiser, A. (2014). *Internal Visuomotor Models for Cognitive Simulation Processes*. Ph.D. thesis, Bielefeld University, Bielefeld.

Kajić, I., Schillaci, G., Bodiroža, S., and Hafner, V. V. (2014). "Learning hand-eye coordination for a humanoid robot using soms," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction, HRI '14* (New York, NY: ACM), 192–193.

Knoblich, G., and Flach, R. (2001). Predicting the effects of actions: interactions of perception and action. *Psychol. Sci.* 12, 467–472. doi:10.1111/1467-9280.00387

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biol. Cybern.* 43, 59–69. doi:10.1007/BF00337288

Kuhl, P. K., and Meltzoff, A. N. (1996). Infant vocalizations in response to speech: vocal imitation and developmental change. *J. Acoust. Soc. Am.* 100(4 Pt 1), 2425–2438. doi:10.1121/1.417951

Kuniyoshi, Y., and Sangawa, S. (2006). Early motor development from partially ordered neural-body dynamics: experiments with a cortico-spinal-musculoskeletal model. *Biol. Cybern.* 95, 589–605. doi:10.1007/s00422-006-0127-z

Lakoff, G. (2014). Mapping the brain's metaphor circuitry: is abstract thought metaphorical thought? *Front. Hum. Neurosci.* 8, 958. doi:10.3389/fnhum.2014.00958

Lakoff, G., and Johnson, M. (1980). *Metaphors We Live by*. Chicago: University of Chicago Press.

Lallee, S., and Dominey, P. F. (2013). Multi-modal convergence maps: from body schema and self-representation to mental imagery. *Adap. Behav.* 1, 1–12.

Lara, B., and Rendon-Mancha, J. M. (2006). "Prediction of undesired situations based on multi-modal representations," in *Electronics, Robotics and Automotive Mechanics Conference, 2006*, Vol. 1 (Cuernavaca: IEEE), 131–136.

Law, J., Lee, M., Hülse, M., and Tomassetti, A. (2011). The infant development timeline and its application to robot shaping. *Adap. Behav.* 19, 335–358. doi:10.1177/1059712311419380

Litman, J. (2005). Curiosity and the pleasures of learning: wanting and liking new information. *Cogn. Emot.* 19, 793–814. doi:10.1080/02699930541000101

Lungarella, M., Metta, G., Pfeifer, R., and Sandini, G. (2003). Developmental robotics: a survey. *Connect. Sci.* 15, 151–190. doi:10.1080/09540090310001655110

Maravita, A., and Iriki, A. (2004). Tools for the body (schema). *Trends Cogn. Sci.* 8, 79–86. doi:10.1016/j.tics.2003.12.008

Maravita, A., Spence, C., and Driver, J. (2003). Multisensory integration and the body schema: close to hand and within reach. *Curr. Biol.* 13, 531–539. doi:10.1016/S0960-9822(03)00449-4

Meltzoff, A. N., and Moore, A. K. (1997). Explaining facial imitation: a theoretical model. *Early Dev. Parent.* 6, 179–192. doi:10.1002/(SICI)1099-0917(199709/12)6:3/4<179::AID-EDP157>3.3.CO;2-I

Metta, G. (2000). *Babyrobot – A Study on Sensori-Motor Development*. Genova: University of Genova.

Möller, R., and Schenck, W. (2008). Bootstrapping cognition from behavior – a computerized thought experiment. *Cogn. Sci.* 32, 504–542. doi:10.1080/03640210802035241

Morse, A. F., Greef, J. D., Belpaeme, T., and Cangelosi, A. (2010). Epigenetic robotics architecture (ERA). *IEEE Trans. Auton. Ment. Dev.* 2, 325–339. doi:10.1109/TAMD.2010.2087020

Moulin-Frier, C., Nguyen, S. M., and Oudeyer, P.-Y. (2013). Self-organization of early vocal development in infants and machines: the role of intrinsic motivation. *Front. Psychol.* 4, 1006. doi:10.3389/fpsyg.2013.01006

Ngo, H., Luciw, M., Förster, A., and Schmidhuber, J. (2013). Confidence-based progress-driven self-generated goals for skill acquisition in developmental robots. *Front. Psychol.* 4:833. doi:10.3389/fpsyg.2013.00833

Nigel, J. T. (2014). "Mental imagery," in *The Stanford Encyclopedia of Philosophy*, Fall 2014 Edn, ed. E. N. Zalta (Stanford: Center for the Study of Language and Information), 1.

Nishide, S., Ogata, T., Tani, J., Komatani, K., and Okuno, H. (2007). "Predicting object dynamics from visual images through active sensing experiences," in *2007 IEEE International Conference on Robotics and Automation* (Roma: IEEE), 2501–2506.

O'Keefe, J., and Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* 3, 317–330. doi:10.1002/hipo.450030307

Oudeyer, P.-Y., Kaplan, F., and Hafner, V. V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Trans. Evol. Comput.* 11, 265–286. doi:10.1109/TEVC.2006.890271

Pape, L., Oddo, C. M., Controzzi, M., Cipriani, C., Förster, A., Carrozza, M. C., et al. (2012). Learning tactile skills through curious exploration. *Front. Neurorobot.* 6:6. doi:10.3389/fnbot.2012.00006

Pezzulo, G. (2007). "Anticipation and future-oriented capabilities in natural and artificial cognition," in *Years of Artificial Intelligence: Essays Dedicated to the 50th Anniversary of Artificial Intelligence* (Berlin, Heidelberg: Springer), 257–270.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., Spivey, M., and McRae, K. (2011). The mechanics of embodiment: a dialogue on embodiment and computational modeling. *Front. Psychol.* 2, 5. doi:10.3389/fpsyg.2011.00005

Pfeifer, R., and Bongard, J. C. (2006). *How the Body Shapes the Way We Think: A New View of Intelligence*. Cambridge, MA: The MIT Press.

Piaget, J. (1954). *The Construction of Reality in the Child*. Ballentine: Ballantine Books.

Pitti, A., Mori, H., Kouzuma, S., and Kuniyoshi, Y. (2009). Contingency perception and agency measure in visuo-motor spiking neural networks. *IEEE Trans. Auton. Ment. Dev.* 1, 86–97. doi:10.1109/TAMD.2009.2021506

Reio, T. G. (2011). "Curiosity," in *Encyclopedia of Child Behavior and Development*, eds S. Goldstein and J. Naglieri (USA: Springer), 453–454.

Reio, T. G., Petrosko, J. M., Wiswell, A. K., and Thongsukmag, J. (2006). The measurement and conceptualization of curiosity. *J. Genet. Psychol.* 167, 117–135. doi:10.3200/GNTP.167.2.117-135

Rochat, P. (1998). Self-perception and action in infancy. *Exp. Brain Res.* 123, 102–109. doi:10.1007/s002210050550

Rochat, P., and Morgan, R. (1998). Two functional orientations of self-exploration in infancy. *Br. J. Dev. Psychol.* 16, 139–154. doi:10.1111/j.2044-835X.1998.tb00914.x

Rolf, M., and Steil, J. J. (2014). Explorative learning of inverse models: a theoretical perspective. *Neurocomputing* 131, 2–14. doi:10.1016/j.neucom.2013.04.050

Roncone, A., Hoffmann, M., Pattacini, U., and Metta, G. (2014). "Automatic kinematic chain calibration using artificial skin: self-touch in the icub humanoid robot," in *2014 IEEE International Conference on Robotics and Automation (ICRA)* (Hong Kong: IEEE), 2305–2312.

Rougier, N., and Boniface, Y. (2011). Dynamic self-organising map. *Neurocomputing* 74, 1840–1847. doi:10.1016/j.neucom.2010.06.034

Saegusa, R., Metta, G., Sandini, G., and Sakka, S. (2008). "Active motor babbling for sensorimotor learning," in *IEEE International Conference on Robotics and Biomimetics (ROBIO 2008)* (Bangkok: IEEE), 794–799.

Schaefer, M., Rothemund, Y., Heinze, H.-J., and Rotte, M. (2004). Short-term plasticity of the primary somatosensory cortex during tool use. *Neuroreport* 15, 1293–1297. doi:10.1097/01.wnr.0000129573.36301.db

Schillaci, G. (2014). *Sensorimotor Learning and Simulation of Experience as a Basis for the Development of Cognition in Robotics*. Ph.D. thesis, Humboldt-Universitaet zu Berlin.

Schillaci, G., and Hafner, V. V. (2011). "Random movement strategies in self-exploration for a humanoid robot," in *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Lausanne: ACM/IEEE), 245–246.

Schillaci, G., Hafner, V. V., and Lara, B. (2012a). "Coupled inverse-forward models for action execution leading to tool-use in a humanoid robot," in *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Boston: ACM/IEEE), 231–232.

Schillaci, G., Lara, B., and Hafner, V. V. (2012b). "Internal simulations for behaviour selection and recognition," in *Human Behavior Understanding, Volume 7559 of Lecture Notes in Computer Science*, eds A. Salah, J. Ruiz-del Solar, Ç. Meriçli, and P. Y. Oudeyer (Berlin, Heidelberg: Springer), 148–160.

Schillaci, G., Hafner, V. V., and Lara, B. (2014). "Online learning of visuo-motor coordination in a humanoid robot. A biologically inspired model," in *2014 Joint IEEE International Conferences on Development and Learning and Epigenetic Robotics (ICDL-Epirob)* (Berlin: Humboldt-Universitaet zu Berlin), 130–136.

Schillaci, G., Hafner, V. V., Lara, B., and Grosjean, M. (2013). "Is that me? Sensorimotor learning and self-other distinction in robotics," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Tokyo: ACM/IEEE), 223–224.

Schmerling, M., Schillaci, G., and Hafner, V. V. (2015). "Goal-directed learning of hand-eye coordination in a humanoid robot," in *Proceedings of the 5th International Conference on Development and Learning and on Epigenetic Robotics*. Providence.

Schrodt, F., Layher, G., Neumann, H., and Butz, M. V. (2015). Embodied learning of a generative neural model for biological motion perception and inference. *Front. Comput. Neurosci.* 9:79. doi:10.3389/fncom.2015.00079

Shaw, P., Law, J., and Lee, M. (2015). "Representations of body schemas for infant robot development," in *2015 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)* (Providence: IEEE), 123–128.

Sposito, A., Bolognini, N., Vallar, G., and Maravita, A. (2012). Extension of perceived arm length following tool-use: clues to plasticity of body metrics. *Neuropsychologia* 50, 2187–2194. doi:10.1016/j.neuropsychologia.2012.05.022

Stoytchev, A. (2005). "Toward learning the binding affordances of objects: a behavior-grounded approach," in *Proceedings of AAAI Symposium on Developmental Robotics* (Stanford, CA: Stanford University), 17–22.

Strack, F., Martin, L., and Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: a nonobtrusive test of the facial feedback hypothesis. *J. Pers. Soc. Psychol.* 54, 768–777. doi:10.1037/0022-3514.54.5.768

Takahashi, K., Ogata, T., Tjandra, H., Murata, S., Arie, H., and Sugano, S. (2014). "Tool-body assimilation model based on body babbling and a neuro-dynamical system for motion generation," in *Artificial Neural Networks and Machine Learning – ICANN 2014, Volume 8681 of Lecture Notes in Computer Science*, eds S. Wermter, C. Weber, W. Duch, T. Honkela, P. Koprinkova-Hristova, S. Magg, et al. (Berlin: Springer International Publishing), 363–370.

Tani, J., Ito, M., and Sugita, Y. (2005). Self-organization of distributedly represented multiple behavior schemata in a mirror system: reviews of robot experiments using rnnpb. *Neural Netw.* 18, 103–104. doi:10.1016/j.neunet.2004.12.001

Thelen, E., and Smith, L. (1996). *A Dynamic Systems Approach to the Development of Cognition and Action. A Bradford Book*. Cambridge, MA: MIT Press.

Udin, S. B., and Fawcett, J. W. (1988). Formation of topographic maps. *Annu. Rev. Neurosci.* 11, 289–327. doi:10.1146/annurev.ne.11.030188.001445

Varela, F. J., Thompson, E. T., and Rosch, E. (1992). *The Embodied Mind: Cognitive Science and Human Experience*, New Edn. Cambridge, MA: The MIT Press.

Von Hofsten, C. (1982). Eye–hand coordination in the newborn. *Dev. Psychol.* 18, 450. doi:10.1037/0012-1649.18.3.450

Watson, J. B. (1913). Psychology as the behaviorist sees it. *Psychol. Rev.* 20, 158–177. doi:10.1037/h0074428

Webb, B. (2004). Neural mechanisms for prediction: do insects have forward models? *Trends Neurosci.* 27, 278–282. doi:10.1016/j.tins.2004.03.004

Weiss, C., Herwig, A., and Schütz-Bosbach, S. (2011). The self in action effects: selective attenuation of self-generated sounds. *Cognition* 121, 207–218. doi:10.1016/j.cognition.2011.06.011

Wexler, M., and Klam, F. (2001). Movement prediction and movement production. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 48–64. doi:10.1037/0096-1523.27.1.48

Wilson, M., and Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychol. Bull.* 131, 460–473. doi:10.1037/0033-2909.131.3.460

Wilson, R. A., and Foglia, L. (2011). "Embodied cognition," in *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta (Stanford: Center for the Study of Language and Information), 1.

Wolpert, D. M., Doya, K., and Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 358, 593–602. doi:10.1098/rstb.2002.1238

Wolpert, D. M., and Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nat. Neurosci.* 3(Suppl.), 1212–1217. doi:10.1038/81497

Wolpert, D. M., Ghahramani, Z., and Flanagan, J. R. (2001). Perspectives and problems in motor learning. *Trends Cogn. Sci.* 5, 487–494. doi:10.1016/S1364-6613(00)01773-3

Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science* 269, 1880. doi:10.1126/science.7569931

Wolpert, D. M., Goodbody, S. J., and Husain, M. (1998). Maintaining internal representations: the role of the human superior parietal lobe. *Nat. Neurosci.* 1, 529–533. doi:10.1038/2245

Yoshikawa, Y., Hosoda, K., and Asada, M. (2004). "Binding tactile and visual sensations via unique association by cross-anchoring between double-touching and self-occlusion," in *Proceedings of the Fourth International Workshop on Epigenetic Robotics* (Genoa: IEEE), 135–138.

Ziemke, T., Jirenhed, D.-A., and Hesslow, G. (2005). Internal simulation of perception: a minimal neuro-robotic model. *Neurocomputing* 68, 85–104. doi:10.1016/j.neucom.2004.12.005

Zoia, S., Blason, L., D'Ottavio, G., Bulgheroni, M., Pezzetta, E., Scabar, A., et al. (2007). Evidence of early development of action planning in the human foetus: a kinematic study. *Exp. Brain Res.* 176, 217–226. doi:10.1007/s00221-006-0607-3

Book review: Contemporary sensorimotor theory

Tom Froese[1]* and Franklenin Sierra[2]

[1] *Departamento de Ciencias de la Computación, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Mexico City, Mexico,* [2] *Instituto de Investigaciones Filosóficas, Universidad Nacional Autónoma de México, Mexico City, Mexico*

A book review on
Contemporary sensorimotor theory

by Bishop, J. M., and Martin, A. O. (eds). (2014). Switzerland: Springer International Publishing

Consciousness, with its irreducible subjective character, was almost exclusively a philosophical topic until relatively recently. Today, however, the problem of explaining the felt quality of experience has also become relevant to science and engineering, including robotics and AI: "What would we have to build into a robot so that it really felt the touch of a finger, the redness of red, or the hurt of a pain?" (O'Regan, 2014, p. 23). Yet a practical response still requires an adequate theory of consciousness, which brings us back to the hard problem: how can we account, from a scientific point of view, for the phenomenological character of experience? Over a decade ago, O'Regan and Noë (2001) proposed a new approach to these questions, the so-called sensorimotor approach to perceptual experience. How far has this approach come and what are its outstanding challenges? The volume *Contemporary Sensorimotor Theory*, edited by Bishop and Martin, takes stock of the current state of the field.

The book starts with Bishop and Martin (2014) presenting different facets of sensorimotor theory, highlighting, for example, that O'Regan (2011) and Noë (2004) ended up developing different ideas concerning the applicability of the theory to robots: a positive account appealing to higher-order cognitive capacities versus a skeptical stance citing the necessity of life for mind, respectively. Ambiguous labeling does not help the current situation. According to Hutto and Myin (2013), the sensorimotor approach of O'Regan and Noë (2001) is also "enactive," a label which Noë (2004) himself began to adopt, but from which Pascal and O'Regan (2008) distanced themselves. In fact, several overlapping approaches may be distinguished in addition to the classic sensorimotor approach, including sensorimotor enactivism (Varela et al., 1991; Noë, 2004), which turned into autopoietic enactivism (Thompson, 2005, 2007; Noë, 2009; Froese and Di Paolo, 2011), and which is distinguished from radical enactivism by Hutto and Myin (2013). The book's contributions range over all of them.

Noë did not contribute to this volume, but his absence is compensated by other submissions. Pepper (2014) points out some conceptual difficulties with Noë's theory of perception, which could be resolved with Merleau-Ponty's phenomenology of the body schema and sedimentation. Wadham (2014) claims that Noë's theory implies the invisibility of perspectival properties, which requires a revision of his theory of perspectival content.

O'Regan (2014) reports on his sensorimotor approach. He proposes that "experiencing a sensation involves being engaged in sensorimotor interaction" but that "being conscious of something [...] requires appeal to a form of 'higher-order' cognitive access" (p. 34). In contrast, Rainey (2014) argues that consciousness is non-conceptual while experience is conceptual, and that consciousness is, therefore, the enabling ground for the possibility of experience.

Scarinzi (2014) points out difficulties faced by O'Regan's approach, characterized as "semi-enactive," that could be resolved by paying closer attention to the lived body, as done by autopoietic enactivism (Thompson, 2007). Paine (2014) also critically examines O'Regan's proposal, evaluating how Heideggerian phenomenology may help his ideas about robot consciousness to evade Dreyfus (2007) objections against AI. Paine also notes that O'Regan leaves out any role for emotion.

This concern is shared by Parthemore (2014), who proposes to extend sensorimotor theory by taking into account emotional affect and the somatosensory system, and to, thereby, turn it into a better theory of concepts. Other authors also propose extensions. Lyon (2014) explores the implications of extending sensorimotor theory beyond vision and touch, in particular to audition. Rucinska (2014) extends sensorimotor theory to explain basic forms of pretense. Cowley (2014) considers how language extends the sensorimotor domain.

There is also an unresolved tension about the role of informational content in the generation of perceptual consciousness in the book. Some authors explore the qualitative differences between types of sensations in terms of information processing (Gamez, 2014), while others advocate abandoning the appeal to informational content altogether (Loughlin, 2014). One problem with a non-representational approach is to explain the experience of imaginary things. Rucinska (2014) account of "seeing-as" may help in developing a solution.

To sum up, this volume invites us to refine our notions of consciousness and experience on the basis of the close relationship between action and perception. However, more work needs to be done to compare and contrast the distinct kinds of sensorimotor/enactive theories. In the context of AI and robotics, for example, we need to clearly distinguish between sensorimotor and autopoietic enactivism. The popularity of the sensorimotor approach is largely explained by its applicability to the design of AI and robotics (e.g., Hoffmann, 2014; Lyon, 2014), and by O'Regan's (2014) claim that it could lead to genuine examples of conscious machines. But this appeal is counterbalanced by a set of philosophical difficulties (Bishop and Martin, 2014), including a lack of clear definitions as to what it means to be an agent or to perform an action (Thompson, 2005).

Autopoietic enactivism, on the other hand, gives us a more solid conceptual foundation of subjectivity by drawing from biological embodiment and from the phenomenological tradition, but not without unfortunate implications for research in AI and robotics (Froese and Ziemke, 2009). Although dynamical systems models of cognition can help us to formally define different notions of sensorimotor contingency (Buhrmann et al., 2013), they are forced to abstract away the autopoietic foundations of agency. Of course, even on this view, research in robotics and the sensorimotor approach continue to form a productive relationship. Yet investigating the hard problem of perceptual experience requires working directly with the first-person perspective. In accordance with the contribution by Gibbs and Devlin (2014), we propose that we can keep the advantages of a synthetic methodology by shifting emphasis from autonomous robotics to human–computer interfaces (Froese et al., 2012). As Gillies and Kleinsmith (2014) propose, such an embodied and enactive approach to designing human–computer interfaces opens up new opportunities for exploring more intuitive interfaces that directly tap into our bodily capacities for perceptual consciousness.

REFERENCES

Bishop, J. M., and Martin, A. O. (2014). "Contemporary sensorimotor theory: a brief introduction," in *Contemporary Sensorimotor Theory*, eds J. M. Bishop and A. O. Martin (Switzerland: Springer), 1–22.

Buhrmann, T., Di Paolo, E. A., and Barandiaran, X. (2013). A dynamical systems account of sensorimotor contingencies. *Front. Psychol.* 4:285. doi:10.3389/fpsyg.2013.00285

Cowley, S. J. (2014). "Human language and sensorimotor contingency," in *Contemporary Sensorimotor Theory*, eds J. M. Bishop and A. O. Martin (Switzerland: Springer), 235–252.

Dreyfus, H. L. (2007). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Philos. Psychol.* 20, 247–268. doi:10.1080/09515080701239510

Froese, T., and Di Paolo, E. A. (2011). The enactive approach: theoretical sketches from cell to society. *Pragmat. Cogn.* 19, 1–36. doi:10.1075/pc.19.1.01fro

Froese, T., Suzuki, K., Ogai, Y., and Ikegami, T. (2012). Using human-computer interfaces to investigate 'mind-as-it-could-be' from the first-person perspective. *Cogn. Comput.* 4, 365–382. doi:10.1007/s12559-012-9153-4

Froese, T., and Ziemke, T. (2009). Enactive artificial intelligence: investigating the systemic organization of life and mind. *Artif. Intell.* 173, 466–500. doi:10.1016/j.artint.2008.12.001

Gamez, D. (2014). "Conscious sensation, conscious perception and sensorimotor theories of consciousness," in *Contemporary Sensorimotor Theory*, eds J. M. Bishop and A. O. Martin (Switzerland: Springer), 159–174.

Gibbs, J. K., and Devlin, K. (2014). "Investigating sensorimotor contingencies in the enactive interface," in *Contemporary Sensorimotor Theory*, eds J. M. Bishop and A. O. Martin (Switzerland: Springer), 189–200.

Gillies, M., and Kleinsmith, A. (2014). "Non-representational interaction design," in *Contemporary Sensorimotor Theory*, eds J. M. Bishop and A. O. Martin (Switzerland: Springer), 201–208.

Hoffmann, M. (2014). "Minimally cognitive robotics: body schema, forward models, and sensorimotor contingencies in a quadruped machine," in *Contemporary Sensorimotor Theory*, eds J. M. Bishop and A. O. Martin (Switzerland: Springer), 209–233.

Hutto, D. D., and Myin, E. (2013). *Radicalizing Enactivism: Basic Minds Without Content*. Cambridge, MA: MIT Press.

Loughlin, V. (2014). "Sensorimotor knowledge and the radical alternative," in *Contemporary Sensorimotor Theory*, eds J. M. Bishop and A. O. Martin (Switzerland: Springer), 105–116.

Lyon, C. (2014). "Beyond vision: extending the scope of a sensorimotor account of perception," in *Contemporary Sensorimotor Theory*, eds J. M. Bishop and A. O. Martin (Switzerland: Springer), 127–136.

Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.

Noë, A. (2009). *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*. New York, NY: Hill and Wang.

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

O'Regan, J. K. (2014). "The explanatory status of the sensorimotor approach to phenomenal consciousness, and its appeal to cognition," in *Contemporary Sensorimotor Theory*, eds J. M. Bishop and A. O. Martin (Switzerland: Springer), 23–35.

O'Regan, J. K., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 883–917. doi:10.1017/S0140525X01000115

Paine, R. (2014). "Heideggerian credentials? O'Regan's sensorimotor approach to perception and robots that feel," in *Contemporary Sensorimotor Theory*, eds J. M. Bishop and A. O. Martin (Switzerland: Springer), 37–52.

Parthemore, J. (2014). "From a sensorimotor to a sensorimotor++ account of embodied conceptual cognition," in *Contemporary Sensorimotor Theory*, eds J. M. Bishop and A. O. Martin (Switzerland: Springer), 137–158.

Pascal, F., and O'Regan, J. K. (2008). Commentary on Mossio and Taraborelli: is the enactive approach really sensorimotor? *Conscious. Cogn.* 17, 1341–1342. doi:10.1016/j.concog.2008.06.013

Pepper, K. (2014). "The phenomenology of sensorimotor understanding," in *Contemporary Sensorimotor Theory*, eds J. M. Bishop and A. O. Martin (Switzerland: Springer), 53–65.

Rainey, S. (2014). "Experience and consciousness: concepts from the outside in," in *Contemporary Sensorimotor Theory*, eds J. M. Bishop and A. O. Martin (Switzerland: Springer), 83–104.

Rucinska, Z. (2014). "Basic pretending as sensorimotor engagement? Lessons from sensorimotor theory for the debate on pretence," in *Contemporary Sensorimotor Theory*, eds J. M. Bishop and A. O. Martin (Switzerland: Springer), 175–187.

Scarinzi, A. (2014). "How enactive is the dynamic sensorimotor account of raw feel? Discussing some insights from phenomenology and the cognitive sciences," in *Contemporary Sensorimotor Theory*, eds J. M. Bishop and A. O. Martin (Switzerland: Springer), 67–81.

Thompson, E. (2005). Sensorimotor subjectivity and the enactive approach to experience. *Phenomenol. Cogn. Sci.* 4, 407–427. doi:10.1007/s11097-005-9003-x

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Harvard University Press.

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Wadham, J. (2014). "The problem of invisible content," in *Contemporary Sensorimotor Theory*, eds J. M. Bishop and A. O. Martin (Switzerland: Springer), 117–126.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read, for greatest visibility

COLLABORATIVE PEER-REVIEW
Designed to be rigorous – yet also collaborative, fair and constructive

FAST PUBLICATION
Average 85 days from submission to publication (across all journals)

COPYRIGHT TO AUTHORS
No limit to article distribution and re-use

TRANSPARENT
Editors and reviewers acknowledged by name on published articles

SUPPORT
By our Swiss-based editorial team

IMPACT METRICS
Advanced metrics track your article's impact

GLOBAL SPREAD
5'100'000+ monthly article views and downloads

LOOP RESEARCH NETWORK
Our network increases readership for your article