Goal-directed tactile exploration for body model learning through self-touch on a humanoid robot

Filipe Gama*, Maksym Shcherban*, Matthias Rolf†, and Matej Hoffmann*

*Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague,
Prague, Czech Republic
filipe.gama@fel.cvut.cz, matej.hoffmann@fel.cvut.cz
†School of Engineering, Computing and Mathematics, Oxford Brookes University
Oxford, United Kingdom
mrolf@brookes.ac.uk

Abstract—An early integration of tactile sensing into motor coordination is the norm in animals, but still a challenge for robots. Tactile exploration through touches on the body gives rise to first body models and bootstraps further development such as reaching competence. Reaching to one's own body requires connections of the tactile and motor space only. Still, the problems of high dimensionality and motor redundancy persist. Through an embodied computational model for the learning of self-touch on a simulated humanoid robot with artificial sensitive skin, we demonstrate that this task can be achieved (i) effectively and (ii) efficiently at scale by employing the computational frameworks for the learning of internal models for reaching: intrinsic motivation and goal babbling. We relate our results to infant studies on spontaneous body exploration as well as reaching to vibrotactile targets on the body. We analyze the reaching configurations of one infant followed weekly between 4 and 18 months of age and derive further requirements for the computational model: accounting for (iii) continuous rather than sporadic touch and (iv) consistent redundancy resolution. Results show the general success of the learning models in the touch domain, but also point out limitations in achieving fully continuous touch.

Index Terms—body exploration, self-touch, goal babbling, intrinsic motivation, reaching development, body schema

I. INTRODUCTION

While recent decades have seen vast progress in robot tactile sensing [1], [2], touch is still rarely used as a primary sense. Research largely focuses on tactile sensing in robot hands for manipulation [3], which is overseen and guided by vision. Systems with the ambition of comprehensive touch as a primary sense are still the exception; if present, touch is an add-on to existing sensing systems (e.g., in the iCub humanoid robot [4]). In stark contrast, touch is the first sense to emerge in the biological fetus [5], and provides a crucial scaffolding by which later motor and cognitive development are grounded [6]. For touch to be used truly as a primary sense, it is necessary to mutually ground and coordinate tactile sensing, vision, and motor control, which poses an enormous challenge to the development of robotic systems. Self-touch may play a crucial role for the development of these skills in infants [7], and indeed seems like a uniquely well-suited calibration scenario: all three modalities are used and create contingent [8] stimuli, which are much less dependent on uncontrollable

external factors than when touching external objects. Generating self-touch requires coordinated motor action, which can be challenging for systems with many degrees of freedom and sparse sensor arrangement. This is true in particular for robots as their tactile sensing is still rather rudimentary and self-contact does not occur naturally like in infants.

The overarching hypothesis of this work is that the similarities and shared challenges between self-touch and reaching for objects external to the body demand for a shared conceptual, technological, and developmental framework. This leads us to the *technological hypothesis* that learning coordinated self-touch can be achieved with models initially proposed for external reaching. Learning to coordinate one's body by means of inverse models [9] that suggest motor actions for desired outcomes is a well-understood concept for external reaching. We seek to transfer knowledge from this well understood domain to the much less understood touch domain, and transfer specific concepts such as goal-directed and structured exploration in order to allow mastery of highly articulated motor systems. Learning inverse models through exploration of random movements–often dubbed *body babbling* [10] or *motor babbling*—has been employed in different models (e.g., the "endogenous random generator" in [11]). However, faced with the dimensionality of the motor and sensory spaces, trying out all possible combinations of motor commands and observing their consequences is hugely inefficient. For example, most motor commands generate movements that do not result in any contact with the body and hence do not generate useful experience to learn the motor-tactile contingencies. Therefore, we employ two key ideas that help the agent to channel the exploration in the right direction. First, the agent should monitor its learning efficiency—the gain in its knowledge or competence to achieve specific goals—and focus the exploration on regions of the search space that are currently most promising. This is exemplified by the computational frameworks dealing with intrinsic motivation (or artificial curiosity) [12]–[15]. Second, the agent should focus the exploration on the goal space (hence "goal babbling") rather than the motor space [14], [16]. The goal space—the skin on the body in our case—may be lower-dimensional and it is where the "interest" of the agent lies. The key challenge that self-touch adds to the learning problem, compared to reaching,

is that most motor actions do not result in any observable self-touch outcome in any region of interest, and that sensible outcomes only lie on a lower dimensional manifold (the skin) in space [17]. In contrast, typical reaching setups [14], [16] involve sensible visual outcomes at least for the majority of possible actions spanning the entire volume of space, which provides much denser and richer feedback.

Complementary to the technological hypothesis, we promote the *developmental hypothesis* that a continuity of mechanisms exists between self-touch and reaching in infant development, rather than distinct mechanisms in each respective stage of development [7]. This aligns with the idea of goal babbling, which itself is a continuity model [18] of exploration and exploitation.

A. Related work

A key requirement for a model to learn self-touch is *efficiency*: the learner has to cope with limited learning time and resources, and learns to coordinate in a potentially high dimensional motor space. Efficient learning of general sensorimotor skills has been studied extensively. Our focus are "mechanisms that drive a learning agent to perform different activities for their own sake, without requiring any external reward" [14]. This phenomenon has been articulated in psychology as intrinsic vs. extrinsic motivation— [19] provides an overview. Oudeyer and Kaplan [13] strive to clarify the terms of internal/intrinsic and external/extrinsic rewards and present a computational perspective as well as the relationship to other computational frameworks such as reinforcement learning. As briefly outlined above, there are two key aspects of efficient exploration: (i) monitoring learning progress and (ii) focusing on the "goal space". The former has been addressed by a number of frameworks that can be classified as *knowledge-based* [13]. The latter aspect has been addressed by the goal babbling approach of Rolf et al. [16] or by other *competence-based* approaches, in which the agent self-generates goals that it tries to accomplish. The idea is best illustrated on the example of learning to reach, or learning inverse kinematics. The motor system is known for its redundancy: there are multiple ways of reaching to a specific point in space. Knowledge-based approaches that monitor learning progress but are confined to the motor space (e.g., see [20]) will discover multiple solutions to the same goal, which can often be considered inefficient. Moreover, the space of solutions in the joint space (motor space) is not convex: averaging between them will often result in wrong configurations. Rolf et al. [16] analyze this and develop a solution, goal babbling, that deals with this problem: by exploring in the goal space, the agent is not "motivated" to look for alternative solutions. Furthermore, following continuous paths through the goal space allows to circumvent the issue of non-convex solutions [16]. This architecture has been also used to model the U-shaped curve typical of infant development [21]. Technologically, it has enabled to address the control of advanced biomimetic robots such as an elephant-trunk-inspired robot [22]. Baranes and Oudeyer extended their R-IAC (Robust Intelligent Adaptive Curiosity) architecture [20] to Self-Adaptive Goal Generation

Robust Intelligent Adaptive Curiosity (SAGG-RIAC) [14]—a competence-based strategy—that also handles learning inverse kinematics in redundant manipulators. Our work is employing the computational framework of [14], as embedded in the *Explauto* library [23]. The algorithms for both modelling and exploration embedded in *Explauto* heavily rely on motor actions being discontinuous in time, whereas path-based approaches [16] rely on continuity and exploit it. Very recent work in [24] has, however, seen a first integration of path-based goal exploration [16] and competence based intrinsic motivation [14].

Learning to discover the surface of the body—a 2-dimensional skin surface embedded in the 3-dimensional world and moving together with the body parts—is similar to the problem of learning inverse kinematics that is a typical show-case for many of the intrinsic motivation frameworks (e.g., see [14], [16]). The motor space or joint space is identical; the goal space, or *observation space*, is different: for learning inverse kinematics, these are the 3D Cartesian coordinates of the end effector (e.g., the infant hand). For the body space, either skin activation or spatial coordinates are candidate representations, which will be explained in detail in Section III. The key difference to reaching in general is that during reaching to one's own body, feedback is not available for many postures. Some, or most (depending on morphology), motor actions will not result in a significant self-touch, in which case there may not be a possible learning step for a model. Whether existing frameworks ([14], [16]) can be *effective* (i.e. functional) under these specific circumstances has not been shown so far.

The work of Kuniyoshi, Mori and colleagues (e.g., see [25], [26]) on the fetus simulator is complementary to this work, addressing prenatal development and focusing on a lower level: first tactile-motor interactions are emerging from the musculoskeletal body model coupled to spinal and simple subcortical or cortical circuitry. In comparison, the present study focuses on how guided exploration on a higher level of abstraction can give rise to efficient body exploration. More importantly, we are looking for goal-directed coordination that will facilitate comparison with infant data from experiments where dynamically chosen locations were designated as touch worthy. The model in [25], [26] does not comprise any inverse models and therefore does not allow for goal-directed movement, but scales by only looking for broad correlations on statically defined interest patterns.

The work most related to ours is that of Mannella et al. [27] who specifically target the body (skin surface) as the exploration target. The biggest experimental difference to the present study is the scale of the task. While we are looking at a humanoid scale motor problem with 2D skin, Mannella et al. only investigate a very simple simulation. It consisted of two arms in 2D with three degrees of freedom (DoF) each, and a "skin" emulated using 30 Gaussian receptive fields in a 1D topology. Their architecture is rather more complex compared to ours, consisting of Goal generator, Goal selector, Motor controller, and Predictor. The motor controller is also highly complex, composed of a dynamic-reservoir recurrent neural network, a random generator, and associative memory. The "skin receptors" are phasic, as they respond

to changes rather than sustained values. These changes are then relayed into a self-organizing neural map (SOM) that "clusters" them. Coordination is achieved by learning motion trajectories through reinforcement. The inner number of variables, or dimensions, is therefore rather high, which makes scaling to humanoid complexity challenging. Compared to this, our architecture is much simpler *and chosen specifically for efficiency* by focusing on direct inverse models that can estimate the necessary posture directly, rather than having to learn to generate entire trajectories. The motor space consists simply of the robot joint space. That is, only the final configurations/postures matter—motor overlaps with proprioceptive—and the actual movement production is sidestepped. Additional discussion and pointers to related work on this issue—learning to reach "synchronously" by matching points in space vs. "asynchronously", taking into account movement trajectories, is presented in [28].

This work is a direct extension of our previous work [29], adding in particular: (i) pilot analysis of infant reaching kinematics for targets on the body (Section II), (ii) path-based goal babbling (after [22]), (iii) improvements of the experimental setup, (iv) new experiments and analyses.

B. Structure and Outline

Section II reviews relevant behavioral studies on infants and presents our own pilot study on reaching configurations used by an infant to reach for targets on the body. We distill two further key requirements to a computational model for the development of self-touch: accounting for not just singular touches, but continuous touching motion, and describing a consistent and repeatable choice of redundancy resolution. Section III presents the robot simulator and the exploration framework. Experimental results (Section IV) show for the first time the effectiveness of existing intrinsic motivation and goal babbling frameworks in the touch domain, as well as their efficiency. We summarize the key findings from the experiments in Section V and discuss their implications and future work in Section VI.

## II. INFANT BODY EXPLORATION AND REACHING FOR TARGETS ON THE BODY

Fetuses initially perform local movements directed to areas of the body most sensitive to touch: the face, but also soles of feet [30, p. 113-114]. Later, from 26 to 28 weeks of gestational age, they also use the back of the hands and touch other body areas like thighs, legs, and knees [30, p. 29-30]. In addition, from 19 weeks, fetuses anticipate hand-to-mouth movements [31] (the mouth opens prior to contact) and from 22 weeks, the movements seem to show the recognizable form of intentional actions, with kinematic patterns that depend on the goal of the action (toward the mouth vs. toward the eyes) [32].

Hand-mouth coordination continues to develop after birth [33]. Specifically related to body exploration, Rochat [34] writes: "By 2-3 months, infants engage in exploration of their own body as it moves and acts in the environment. They babble and touch their own body, attracted and actively involved in investigating the rich intermodal redundancies, temporal contingencies, and spatial congruence of self-perception." Thomas et al. [35], biweekly recording resting alert infants from birth to 6 months of age, show that infants frequently touch their bodies, with a rostro-caudal progression as they grow older: head and trunk contacts are more frequent in the beginning, followed by more caudal body locations including hips, then legs, and eventually the feet. DiMercurio et al. [36], following infants from 3 to 9 weeks after birth, found no consistent differences in the rate of touch between the head and the trunk. In summary, infants acquire ample experience of touching their body, which allows for the learning of the first tactile-proprioceptive-motor models of the body. The ability to learn from this experience goes hand in hand with dynamic neural development in this period [37]; see [7] for a review focusing specifically on self-touch. Yet, the behavioral organization of such early tactile exploration not understood. Are the touches on the body spontaneous or systematic? If there is a particular structure—which seems to be the case [35], [36]—what drives this developmental progression? Piaget [38] theorized that in newborns, action and perception as well as the "spaces" of individual sensory modalities are separated (cf. [39] for evidence that visual and motor modalities are connected early after birth). Until the connections are established, infants explore their environment (and their body) randomly. Piaget [38] also proposed a pivotal role of repeated movements—*primary circular reactions* directed to learn properties of the body and *secondary circular reactions* driven by the interest on the effects they produce in the environment. A computational account combining rhythmic and discrete movements and reinforcement learning is presented in [40].

Behavioral studies investigating infant spontaneous behaviors with a specific focus on touch to the body [35], [36] provide data that inform our modeling. First, in the first weeks after birth, contacts with the rostral areas of the body (head and trunk) are dominant. Second, contacts are typically made with the ipsilateral hand. DiMercurio et al. [36] also identified *complex touches* (as infants moved their hand while remaining in contact with their body), performed a network analysis of contact sequences, and identified points of centrality on the body. However, to discriminate spontaneous contacts from systematic (intrinsically motivated) exploration remains a challenge. The observation of frequent "complex touches" raises an important requirement for a computational model—it has to explain the *continuous* coordination of the task across isolated touch points. Purely conceptually, path-based exploration approaches [18] seem to fit the requirement of exploration "along the skin surface" more naturally than sporadically sampling ones [14], where every movement is started anew from some canonical ("home") posture. However, sporadic exploration at some point also has to move between points, even though that is typically abstracted away and only end points are considered. Which approach creates experimental data closer to infant observations is not clear yet, partially pending further quantitative analysis of infant data.

A counterpart to recordings of spontaneous infant behavior is provided by testing how they can reach to targets on their body. Lockman and colleagues performed a series of
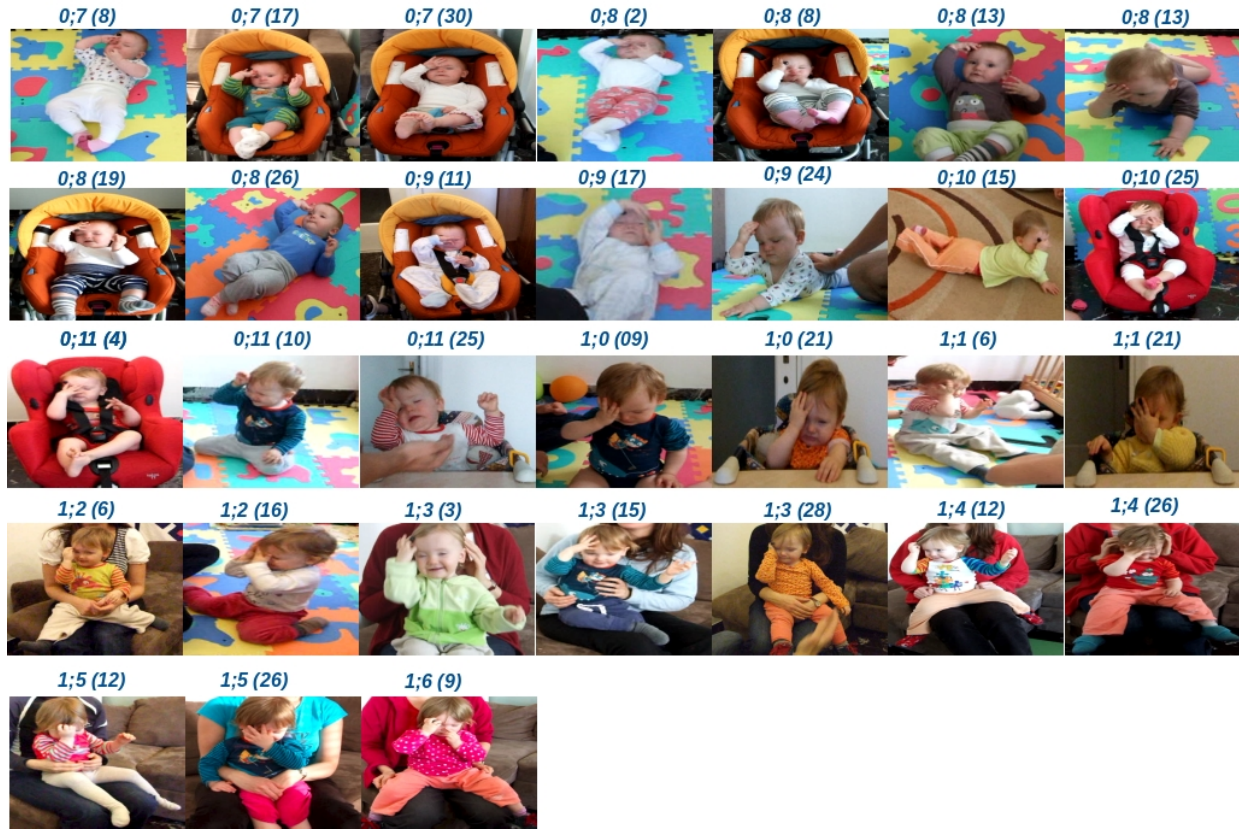
(a) Right forehead

(b) Left trunk

Fig. 1: Infant reaching to the body. Screenshots of final postures while reaching for a buzzer on: (a) Right forehead, (b) Left trunk. One infant followed from 4 to 18 months. Only screenshots from sessions where she successfully contacted the target are shown. Age is reported in the notation years;months (days).

studies [41]–[45] in which vibrotactile targets ("buzzers") were attached to infants' body parts and their ability and their way of reaching for the targets were analyzed. Targets above the mouth and on the chin were successfully contacted already from 2 months of age [42], followed by trunk area, legs, hands, other areas on the face (forehead, ears), and elbows

(around 9 months) (whole body – pilot study [43]; upper body [44]). For targets on hands and arms, the arm with the buzzer and the contralateral arm reaching for the target often moved simultaneously—the arm with the target actively facilitating the removal [41].

One aspect remained unexplored so far: how infants manage motor redundancy in this task. Do they use the same arm configuration to reach for specific targets on the body or do they have alternatives at their disposal? If the latter is true, what does the choice of solution depend on? Switching solutions may depend on the initial/current posture— what is known as motor hysteresis [46]. Finally, what is the developmental progression of this phenomenon? This kind of information is important for the modeling work, as it constraints the inverse model—a mapping from body space to joint space—representation. To this end, we performed an analysis of reaching configurations of one infant tested weekly with buzzers on her body, between 4 and 18 months (experimental protocol described in [44] in detail; summary of success in buzzer removal for this infant reported in [43]; study approved by Tulane University Social-Behavioral IRB). From video recordings of the experiments, snapshots of final postures—around the first contact with the target—were taken. In this article, reaching to the face and trunk will be studied. Hence, we show the final postures the infant assumed to reach for the right forehead (Fig. 1, top) and the left trunk (Fig. 1, bottom). Interestingly, despite the fact that the overall posture of the infant differed significantly (supine, seated in a seat / on the lap, on the tummy, freely sitting) and that the snapshots span more than one year of the infant's life, the final reaching configurations are largely similar. Thus, the constraints on the inverse model can be relaxed. Although it seems clear that, at least eventually, alternative reaching configurations will be available [46], they are not frequently spontaneously recruited. This observation motivates the fourth key requirement: the model needs to describe a *consistent* resolution of redundancy, rather than selecting different postures in every attempt or across trials. This is consistent with the notion of a direct inverse model which stores exactly one solution to reach for any goal. However, consistent behavior may be distally observable also for different internal organizations of the motor control skill.

III. Embodied computational model – Materials and Methods

This section provides an overview of the robot simulator and the exploration framework.

A. Nao humanoid robot with artificial skin

The experimental platform of this study is a Nao humanoid robot. A modified version, uniquely equipped with artificial sensitive skin (Fig. 2), is available at our institute. The electronic skin covers the robot's wrists, torso, and head (Fig. 2, right). Every triangular module hosts ten pressure-sensitive elements, or "taxels" [47]. The physical robot is currently not employed in the results shown here—long exploration experiments would be damaging to the hardware—but the simulated version closely mimics it to allow transition to the real robot in the future.
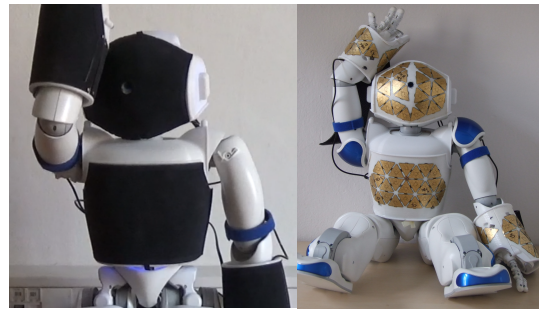

Fig. 2: (Left) Nao robot with skin performing self-touch. (Right) Robot with exposed artificial skin.

Gazebo 9 simulation environment was used. A variant of the publicly available *naov40* URDF model was additionally fitted with tactile/pressure sensors ("skin"), mimicking our physical Nao robot (Fig. 2). In addition, a cylindrical "pen" tool with a spherical endpoint was attached to the robot wrist to act as a finger and facilitate localized touch. The model has 250 and 240 tactile sensors for the torso and the head, respectively. See Fig. 3. Compared to [29], additional modifications were performed to the simulator that included simplification of the meshes composing the model, buffering the contact events, and disabling collisions for other parts of the arm than the finger. Overall, these changes resulted in faster execution and more stable results. The code of our `nao-gazebo-skin` plugin is available at [48]. Videos from our experiments are available at https://youtu.be/dnJaffBHf1c.
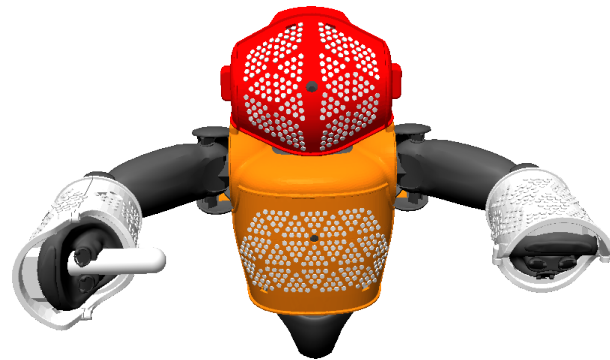

Fig. 3: Nao robot model in Gazebo with tactile sensors.

B. Action and observation spaces

The *action space Q* consists of the robot joint angles. An action $q \in Q$ generates an outcome $x \in X$ in the *observation space X*. Only the upper body of the Nao robot, which hosts the artificial skin, is used. The robot uses its right arm to touch either the torso or the head. Its *action space* is the robot's joint space, with five Degrees of Freedom (DoF) per arm and two DoF on the neck. To touch the torso, only the arm is used, hence $Q \subseteq \mathbb{R}^5$; to touch the head, the neck joints may also contribute: $Q \subseteq \mathbb{R}^7$. The effect of these two joints is specifically studied by comparing with a configuration without

the head joints (*nohj*). Position control in every joint is used to command the simulated robot to a desired joint configuration.

The *observation space* is the robot skin activation generated when the robot contacts its torso or face with its arm. This is a discrete space of individual taxels and their activation (binary: activated or not). For the exploration methods considered here, a distance *metric* on this space is needed. A parallel planar projection (Fig. 4) is used (differently from [29] where a cylindrical projection was used for "high-resolution" skin). Thus, for the torso and head skin, $X \subseteq \mathbb{R}^2$.
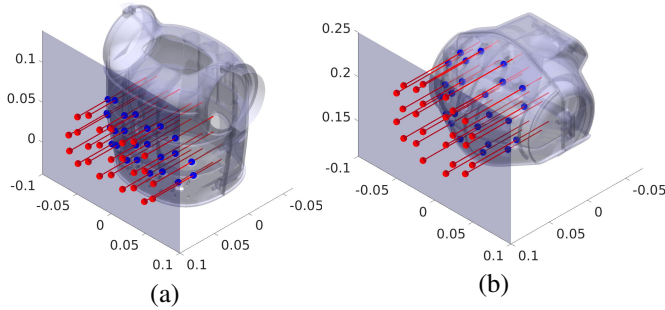

Fig. 4: Parallel planar projection of taxels' coordinates for (a) the torso and (b) the head.

C. Touch vs. Motor and Task Errors

Not all actions or postures in this scenario result in a measurable touch outcome. Fig. 5 illustrates various cases for the simplified scenario of a 2 DoF planar arm, but for which the action space can be fully plotted. A touch signal is generated if the tip of the arm gets in contact with the surface of the object. The observation space of one-dimensional touches is color coded for different touch coordinates x together with actions that achieve precisely that touch outcome (shown both as postures in the physical space, and action coordinates in Q; note that the physical space is two-dimensional while the observation space is only unidimensional—contour of the grey object). Some postures, such as (1), do not generate a touch. Others, (2), would penetrate the body and are impossible (the entire grayed out area in the action space). Some postures generate physical contact, (3), but may not be registered as a touch sensation if performed with an area without sensors, such as the elbow.

Fig. 5, bottom, shows different errors and deviation that can occur when performing a goal directed touching motion. The robot in this example could start from a neutral reset posture. Trying to reach for some touch goal x^*, the inverse model suggests a posture q^*. The underlying motor controller is tasked with moving the robot to q^*, which is physically unreachable in this example. Instead, the arm at some point collides with the skin and is potentially dragged along it while the motor controller is still trying to reach q^*. The robot eventually ends up in a different posture q, with a touch coordinate x (but may generate a non-touch event in other cases). When attempting an impossible action q^*, a motor discrepancy—difference between the desired joint configuration q^* and actual configuration at the end of the
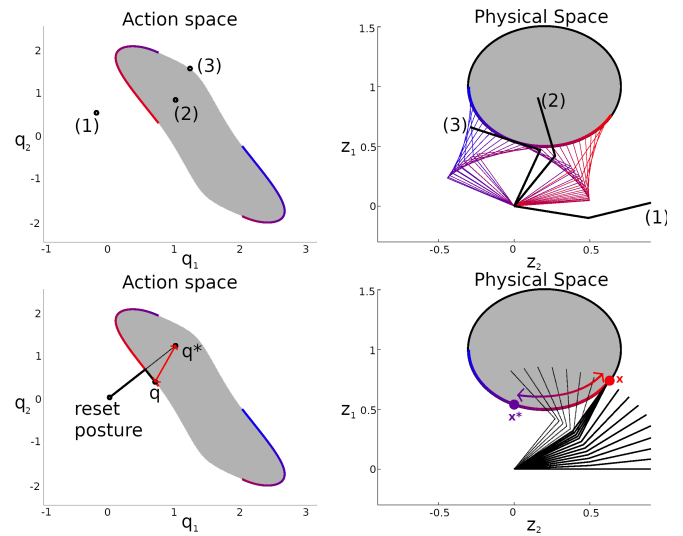

Fig. 5: Exemplary relation between the action space of a 2 DoF arm, and its potential touch on the one-dimensional boundary (observation space) of a body. Top: Examples of different types of touch (color coded) and non-touch events (numbered) – a non-touching posture (1), an impossible posture (2), and a posture that generates a contact with the object that does not register as touch (3). Bottom: exemplary reaching attempt to a rest or reset posture with both a reaching error in observation space, and a motor discrepancy.

movement q—is inevitable. A different kind of error is to miss the reaching target x^* by x. This particular motion is an example of a single *sporadic* touch. Complex, or continuous, touches would describe a motion between several different goals x^* while maintaining touch with the skin. During such complex motions, the motor discrepancies can be subject to hysteresis as the effector is dragged along the skin.

D. Home posture

The default "home" posture of the model in Gazebo that is used to start reaching movements is similar to the one shown in Fig. 3, with the difference that the wrist is rotated such that the "finger" points downward. For exploration methods relying on separate movements to collect every action-observation pair—referred to as *sporadic touch* here—this posture is of little importance, provided that this posture does not impede reaching to the goal space (torso or head) due to self-collisions. However, for path-based continuous approaches (Section III-F), the home posture is an important part of the exploration process and it is beneficial if it lies in the goal space. To this end and in order not to introduce biases or asymmetries, home postures with the hand touching the central part of the torso/head were chosen—see Fig. 6. This home posture is used both for Path-Based Continuous Goal Babbling and as the base posture to bootstrap the Explauto methods.

E. Explauto autonomous exploration library

Explauto (https://github.com/flowersteam/explauto, [23]) is a framework for implementation and benchmarking of sensorimotor learning algorithms, with a specific focus on intrinsic
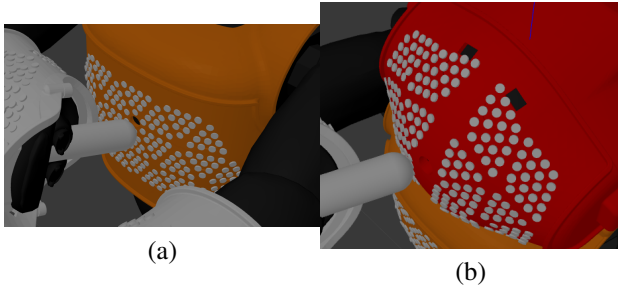
Fig. 6: Home posture providing a touch feedback for (a) the torso and (b) the head.

motivation—monitoring learning progress in motor or sensory (goal) spaces. For a detailed and formal explanation, the reader is kindly referred to [14], [23], [49], [50]. During exploration, a database is constructed with every entry being a tuple: (q, x). We use the *nearest neighbor* model from Explauto to represent both the forward and inverse model (together, the sensorimotor model in the terminology of Explauto). This is an example of *lazy learning*: training data is processed only when a query is asked. Our focus is on inverse models: learning how to reach for particular locations on the skin ($\sim$ inverse kinematics). Given an observation x^*, the inverse model will return the joint configuration q^* that corresponds to the observation stored in the database that is closest to x^*. Forward models are needed only for the Discretized Motor Babbling exploration strategy (see below) that uses forward predictions to gauge the learning progress.

To choose where to focus exploration, an *interest model* is needed, sampling from the *interest space*, which can be either the action space Q (motor babbling strategies) or the observation space X (goal babbling strategies). The exploration strategies from Explauto used in this work are:

a) Random Motor Babbling (RMB): a motor configuration $q^* \in Q$ is sampled uniformly from the action space, and then executed, generating an observation $x \in X$.

b) Random Goal Babbling (RGB): a goal x^* is sampled uniformly from the observation space, and the inverse model is used to find an action q^* best matching the goal, with added exploration noise.

c) Discretized Motor Babbling (DMB), (or Motor Babbling with Intrinsic Motivation): The interest space—the action space—is discretized into $c = m \times n$ cells (regions). We use 32x32 cells. The algorithm randomly selects one of the cells with a probability proportional to the current state of an *interest value* I that each cell possesses. A motor command q_c^* within that cell is generated. The forward model is used to make a prediction of the observation, x_c^*. The motor command is executed, resulting in the real observation x and allowing for calculation of an error (expected vs. real). The interest value of a cell is high when *competence* $C = ||x_c^* - x||$, rapidly increases or declines. The *local competence progress* is formally defined in [49].

d) Discretized Goal Babbling (DGB): This method is analogous to DMB but with the interest space being the observation space (the skin) and the competence progress gauged on the performance of the inverse model. The goal generation randomly selects one of the cells, proportionally to their interest value I. Then, a goal x_c^* within that cell is uniformly generated. The robot attempts to reach for x_c^* using the inverse model (with exploration noise).

All experiments using *Explauto* were run for 1000 iterations. RGB, DMB and DGB require a bootstrapping phase, during which a few touches have to be generated. This phase is counted toward the 1000 iterations limit. To this end, RMB with constrained joints range is used by taking the *home posture* as a base and adding random exploration noise that fits within the constrained range until 10 touches are observed. While a single touch was enough to bootstrap these methods, the use of additional bootstrapping touches led to more consistent results between trials of the same experiment (with small improvements from 5 to 10 touches and no effects with additional ones).

F. Path-Based Continuous Goal Babbling (PBCGB)

In addition to *Explauto* and related SAGG-RIAC methods, we employ the path-based continuous goal babbling method originally developed in [16] with modifications previously employed on the biomimetic elephant trunk robot [22]. While *Explauto* explores by means of sporadic random motion and discontinuous goal selection, [22] samples motor action along continuous paths. A goal x_E^* is chosen randomly from the overall set of goals as the end point of a motion. From the previous end point x_S^*, a straight path of sub-goals $x^*(t)$ is then sampled toward x_E^*. Each sub-goal $x^*(t)$ along the path is attempted to be reached by querying the inverse model for a motor command q_t^*. A small random motor perturbation $\epsilon(t)$ is added to the command to encourage exploration. In order to ensure continuity, this perturbation is not chosen independently in every time step, but generated by an autoregressive random process similar to a Brownian motion that can explore over a wide range of values by accumulating small random steps (see [16], [22] for details). The perturbed motor action $q^*(t) + \epsilon(t)$ is then executed on the robot in order to try and reach for $x^*(t)$ along the continuous goal path. The resulting observation $x(t)$ is then used together with the motor command $q^*(t) + \epsilon(t)$ to train the inverse model. After attempting to reach the end point and each sub-goal in its path, either another end point is selected, or, with a fixed probability, the *home posture* is designated as the next end point.

This is illustrated in Fig. 7: we start from the home posture (red dot in the center, 29), generate the first end point (red dot 7), with sub-goals between it and the home posture (blue dots from 1 to 6). The model attempts to reach for 1, then 2, until 7. After the attempt to reach for 7, a new end goal is generated (red dot 15), with its sub-goals. This happens again to reach for 22, at which point the home posture was chosen instead of another end goal. When the algorithm goes back to the home posture, it uses the difference between the current position and the home posture to reach it in steps—here, 7 steps from 22 to 29—but without generating sub-goals. This ensures that the robot is back on the home posture to start a new exploration path no matter where it was, as there is no guarantee that the goals were successfully reached.
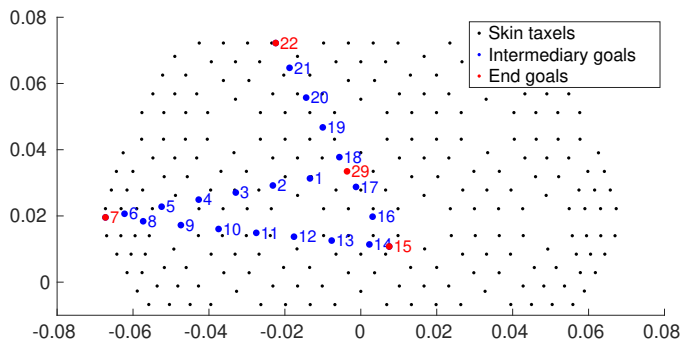
Fig. 7: Example of exploration paths and goals generated by PBCGB. End points in red, sub-goals in blue, taxels in black.

An advantage of this continuous strategy is that the way redundancy is resolved by the eventual inverse model is typically very smooth and does not involve unnecessary jumps. It can moreover be controlled by choosing a home posture, from which exploration starts and to which it repeatedly returns. If, for example, the home posture is chosen in an "elbow up" configuration, the path based approach will develop an inverse model that stays close to it and moves continuously around it to reach goals. SAGG-RIAC [14] cannot achieve this continuity because its sporadic random actions will sample "elbow up" and "elbow down" postures equally, and the Nearest Neighbor model has no means to reconcile them smoothly. Hence, the inverse model would suggest repeatedly switching between both kinds of redundancy resolutions while trying to generate a continuous motion.

An important implication of focusing on a single, consistent way of resolving redundancy is that this strategy has been shown to work in extremely high-dimensional motor spaces, such as 50 [16] or 100 [51] motor dimensions. The main drawback is that the creation of continuous paths and explorative perturbations requires more parameters (e.g., step widths and amplitudes) than the sporadic exploration of SAGG-RIAC which only requires to know the outer size of the goal and motor spaces. The path based approach is therefore more difficult to use "off the shelf". However, the path parameters typically allow for effective experimentation over a wide range of values [16]. The parameters used here were:

- discretization of the observation space, generating the main end goals (not sub-goals): 10x10 list of end goals, uniformly spread on the space.
- probability to go back to the home posture instead of reaching for another end goal: 10%.
- maximum distance between two successive sub-goals, in meters: 0.01.
- maximum distance joints can move in a single step, in radians: 0.1.
- motor perturbation: maximum amplitude of added noise that can influence a joint's position, in radians: 0.01.

G. Learning and testing models

In each experiment, the robot uses a given exploration strategy for a certain number of steps: 1000 for *Explauto* methods,

and 15000 for Path-Based Continuous Goal Babbling. Unlike standard cases in which every iteration of active exploration results in reaching a point in the observation space and allows for calculating an error (target vs. actual outcome—like in the case of learning to reach using vision), in our case, the movement does not always result in contacting the skin. In that case, the step is counted toward the maximum number of iterations but does not contribute to learning the sensorimotor (forward or inverse) model.

At each step, the target joint configuration (motor command sent to the robot) and the actual joint configuration are compared. If the difference for any joint is higher than 10% of its maximum range, the results are discarded and are not used to learn the model. The main reason for such discrepancy is self-collision. Motor commands that would end up "inside the robot" should be avoided. Because of collisions, such commands often make the finger slide on the skin, creating an important difference—sometimes equivalent to the distance between two or three taxels—between the initial contact point and the end point after sliding. This sliding behavior should be avoided as it can influence the results in several ways. For example, an initial contact that does not provide tactile feedback could—if the sliding finger activates a taxel a short time after—be recorded and counted as successfully reaching for the activated taxel.

A subset of taxels, shown in Fig. 8, is tested every 100 iterations during the first 1000 iterations; then every 1000 iterations up to 15000 in case of PBCGB, whereby testing refers to the robot being asked to reach for each test taxel using the current inverse model. In the same manner as for the exploration phase, if no taxel is contacted, no error can be measured. For all experiments, the environment is set to reset at each iteration during the testing phase, and set back to its previous reset settings before exploration starts again.
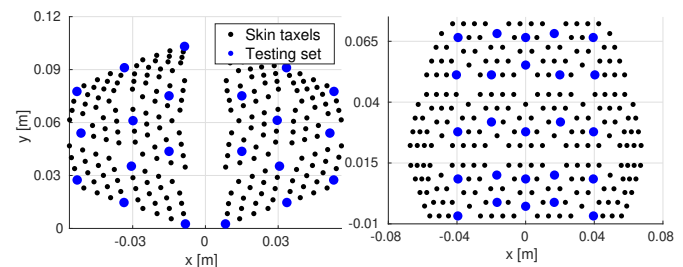

Fig. 8: Testing set. Head (left); Torso (right). All skin taxels in black; testing set in blue.

IV. EXPERIMENTS AND RESULTS

We present the results of a series of experiments. The right hand is the robot's effector and reaches either for the torso skin (Section IV-B) or for the face skin (Section IV-C). In the latter case, the action space will be larger as two neck joints are available in addition to the five arm joints. Videos illustrating our experiments are available at https://youtu.be/dnJaffBHf1c.

We will illustrate the results in the following ways: (i) Mean Reaching Error (MRE) every 100 iterations until 1000, then every 1000 iterations (for PBCGB) (e.g., see Fig. 12, left),

(ii) number of touch occurrences during exploration every 100 iterations until 1000, then every 1000 iterations (e.g. see Fig. 12, right), (iii) projection of the generated goals with details about the reaching error for each test taxel after 1000 (or 15000) iterations (e.g., see Fig. 9). The results are averaged over ten trials for each exploration strategy; any result specific to one trial or averaged over a different amount of trials will be clearly stated. For projections, the observation space is presented from the point of view of an observer looking at the robot—like in Figs. 2 and 3.

For the experiments on the torso, the *Explauto* exploration methods (Sec. III-E) and the Path-Based approach (Sec. III-F) are complemented by a special version of DGB with no reset, *DGB-nr*, whereby the environment is not reset at each iteration and thus the robot starts from the previous configuration— partially mimicking "complex touches" and the operation of PBCGB.

For the head, DGB, PBCGB and *DGB-nr* were run. In addition, two more experiments were added: first, a version of DGB, *DGB-nohj* ("no head joints") for which the neck joints were disabled. Second, a version of PBCGB, *PBCGB-incgs* ("increased goal step") with a modified discretization (9x9) and maximum distance between two successive sub-goals (0.025). The significance of the parameters necessary to organize paths in PBCGB has not so far been exhaustively explored, which gives rise to the expectation that further tuning could enhance its performance. When using the same parameters used for the torso, we observed many sub-goals being created in the central area of the head, where there are no taxels. We decided to include both variants here to compare their results.

The difference between RMB, RGB and DGB in [29] and this work are (1) the projection of the observation space, impacting the calculation of the reaching errors, (2) the discarding of target motor commands that are too different from the actual motor configuration and (3) changes to the simulator described in Section III-A.

A. Analysis of motor discrepancies

Unlike for reaching in free space, reaching to the body inevitably results in self-collisions. Interesting preliminary results arose related to the comparison between target and actual joint configurations and the discrepancies between them. To avoid the sliding effect (Section III-G) and improve the consistency between what the models learn and what the robot actually does, we tested feeding the models with the actual joint configuration after a command instead of the target command. Over five trials, we observed a significant drop of performance and compared the results with another variant where we used the target motor command by default and replaced it with the actual joint command when the discrepancy for at least one of the joints was 10% higher than its absolute range of possible values. The difference in performance was barely noticeable for the torso, but striking for the head, as shown in Fig. 9, as a significant amount of taxels are unreached when using the actual joints, and reached with small errors when using the target joints.
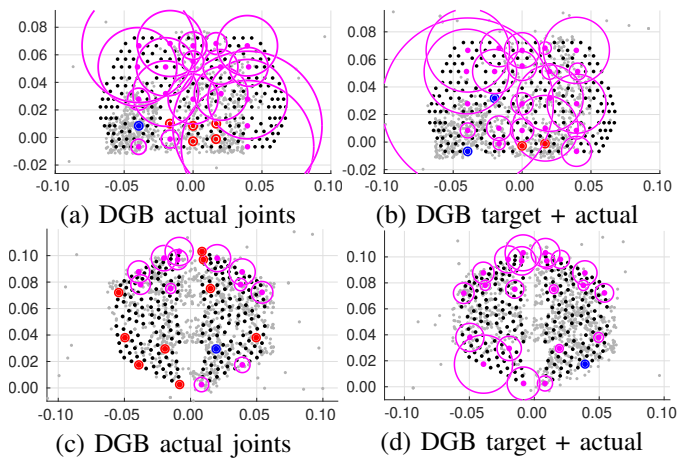

Fig. 9: Right hand reaching for torso (a and b) and head (c and d) – observation space. Comparing the mean reaching performance over 5 trials between using the actual joint configuration for learning (left) and the target joint configuration (right). *Goals* generated during the exploration process (grey). Testing after all learning iterations: *Reached taxels* with no error (blue); *Reached with error* – taxels reached (magenta dots) with error magnitude (magenta circles); *Unreached* taxels – no taxel reached during reaching attempt (red).

The difference is not only present in the reaching performance but also, maybe more importantly, in the number of touches generated: for DGB (both head and torso), there is a +50% number of touches between the two cases, on average 400 touches when using the actual joints, and slightly above 600 for the mixed target and actual joints version, as seen in Fig. 10. These results were observed not only for DGB, but also for PBCGB (though only with 2 trials)—both for the torso and the head. For PBCGB, using mainly the target motor command provided about twice as many touches than using only the actual joints after 15000 iterations, but the variance between trials was higher.
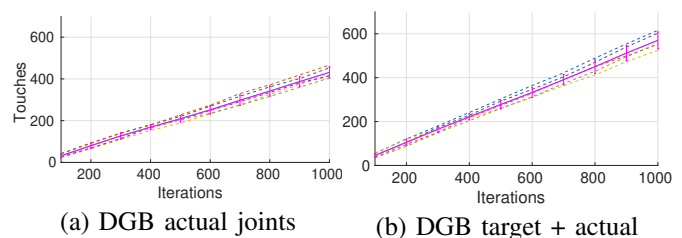

Fig. 10: Right hand reaching for the torso. Touches generated during exploration using the actual joint configuration (left) and the target joint configuration (right). Each dashed line correspond to a trial. The full line corresponds to the mean, with error bars corresponding to the standard deviation.

Thus, commands that try to "penetrate inside the robot" are important to ensure more consistent tactile feedback, as these commands typically exert more force by trying to reach an impossible position, contrary to commands using the actual joint configuration that may end up only brushing against the skin, and not applying enough pressure. To keep the

advantage of these commands while reducing the occurrence of the sliding effect and its impact on the results, the solution described in Section III-G was chosen, where the target motor configuration is used but entirely discarded (instead of being replaced by the actual configuration as in the results above) if the difference with the actual configuration is above the 10% threshold for any joint.

To further analyse motor discrepancies, we used histograms of the errors (difference) between target and actual configurations. Each joint was analysed separately to estimate its range of errors. We also separated the commands that resulted in a tactile feedback, including the discarded ones shown outside the threshold (Fig. 11, left), from the commands that did not provide any feedback (Fig. 11, right).
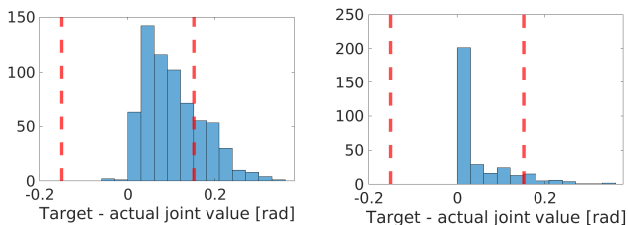

Fig. 11: Joint error through all iterations of the Elbow Roll joint in a single trial from DGB when reaching for the torso. The red vertical lines are the threshold. Iterations with tactile feedback (left), and without (right).

The error ranges for the commands that did not provide tactile feedback mainly stayed within the 10% threshold (see Fig. 11, right). For the majority of iterations, there is zero, or a very small, error. This was expected as commands that do not produce a tactile feedback are often not in contact at all with the robot. Without collisions, there is little discrepancy between the target and actual commands. For commands that produced tactile feedback, the range of errors is more spread (see Fig. 11, left).

Over all methods and trials, the joint that was more often outside the threshold and participated in the discarding of the target command is the Elbow Roll, followed by the Shoulder Pitch, Shoulder Roll, and the Head Pitch. The Elbow Yaw and Head Yaw had the narrowest ranges of errors, keeping an error close to zero and rarely going over the threshold.

B. Reaching for the torso

For each method, we look at the average over all trials of: the Mean Reaching Error and the number of touches through exploration (Fig. 12), and the projection of the skin alongside the distribution of goals generated and the reaching error for each tested taxel (Fig. 13).

Surprisingly, RGB has the lowest MRE after 1000 iterations, close to DGB and DGB-nr. Continuous motion attempt with DGB-nr has very close MRE to DGB with its sporadic motions, contrary to PBCGB that shows a MRE about thrice the value of DGB (Fig. 12, left). DMB displays a low amount of touches, close to RMB, while PBCGB is above RGB but far below both DGB variants which have the highest amount of touches (Fig. 12, right). Looking at the projections (Fig. 13)

reinforces the idea that MRE and the number of touches are not good enough metrics by themselves: while somewhat in the middle in MRE and touches, PBCGB shows a high amount of completely unreached taxels, compared to DGB and DGB-nr (Fig. 13, d and e), and surprisingly does not show major improvements even after 15000 iterations (Fig. 13, f and g). The difference between DGB and RGB (Fig. 13, d and c) is small despite the gap in the number of touches. Compared to [29], DGB shows a smaller amount of touches and of consistently perfectly reached taxels. DGB-nr has similar or higher reaching errors than DGB, and no taxel is consistently perfectly reached, but has several taxels with mean errors smaller or equal to the distance from the target taxel to its closest neighbor.

PBCGB takes on average 6 steps to go back to the home posture and 6.5 steps toward an end goal (i.e., there are on average 6.5 goals between two end goals), has 1895 end goals for 15000 iterations, and it goes back to the home posture 206 times. At 15000 iterations, it averages 4777 ± 1078 touches.
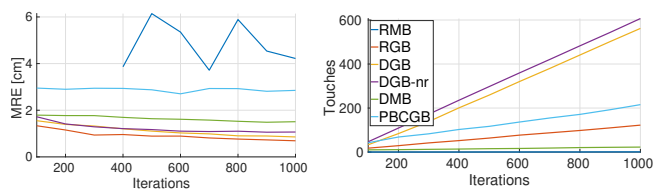

Fig. 12: Right hand reaching for torso – comparison of exploration strategies at 1000 iterations. (Left) MRE. (Right) Number of touches.

For the torso, we also retrieved the 3D errors between the end effector and the tested taxels. These errors serve as an external evaluation for unreached taxels during testing (in red in the projections)—they are not used for learning. The 3D errors are less accurate than the 2D errors, as shown in Fig. 14 for DGB: reached taxels (in blue) will have an error due to the physical size of the objects even if the finger is in contact with the target taxel, or due to small movements (e.g. sliding) that can happen in the delay between the touch and the recording of coordinates. The 3D errors in PBCGB are quite high. They seem to intersect around the taxel touched from the home posture, but an analysis of screenshots of the reaching attempts showed that the finger is often away from the skin, rather than around the home posture.

The mean percentage of discarded motor commands is given in Table. I.

RMB	RGB	DGB	DGB-nr	DMB	PBCGB
0.3±0.1	25.8±9.2	18.6±9.5	17.7±6.1	6.0±5.2	2.4±3.9

TABLE I: Mean percentage of discarded motor commands for torso reaching, relative to 1000 iterations, except PBCGB at 15000 iterations.

C. Reaching for the head

DGB-nohj performs almost as well as DGB with the neck joints enabled with regard to MRE and the number of touches (Fig. 15). The biggest difference is in DGB reaching more
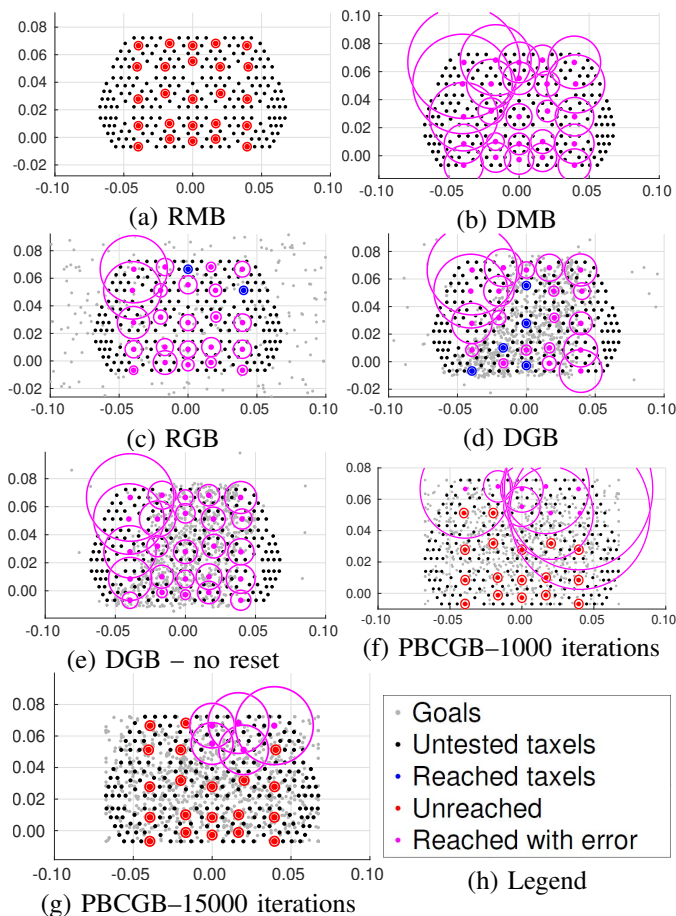
Fig. 13: Right hand reaching for the torso – observation space. See Fig. 9 for details.
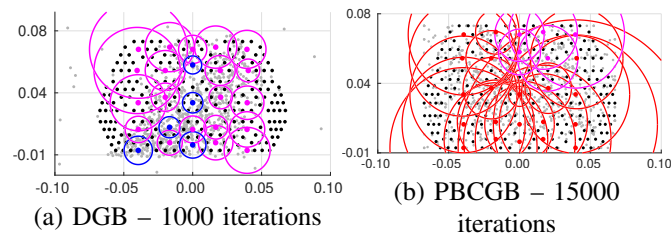

Fig. 14: Right hand reaching for the torso – observation space – using 3D errors. See Fig. 9 and 13 for details.

accurately the ipsilateral side, while DGB-nohj has higher errors there and instead reaches accurately the opposite side of the face (Fig. 16, a and c). The goal generation also highlights this shift.

DGB-nr shows a clear increase of the reaching error in the projections (Fig. 16, b), despite keeping up with the number of touches and a MRE curve that seems close to DGB (Fig. 15). Again, the goal generation is impacted, and more goals are selected on the middle part of the face with no taxels.

PBCGB only achieves poor performance on the head after 1000 iterations with the default version; no taxel was reached during testing and no error could be measured (Fig. 16, d). The version with the increased step length reaches, but with a high error—the intersection of the circles seems to

indicate that it mostly reaches for taxels very close to the home posture (Fig. 16, e). The number of touches is low for both versions (Fig. 15, right). After 15000 iterations, the performance improves significantly. There were on average 4778±626 and 5161±402 touches and the MRE dropped to 1 cm. The projections (Fig. 16, f and g) show that reaching for the upper-right part of the head was learned, but the remaining part of the skin is not even reached with error. The goals projected are also mainly present in the center of the middle stripe of the face. This is mainly due to how the algorithm functions, as these include the sub-goals, and the middle area is bound to be covered by sub-goals when attempting to reach from one side of the face to the other.

We let one of the trials for PBCGB run until 30000 iterations (Fig. 16, h) instead of stopping at 15000. Its projection shows that the whole upper part of the head is reached quite accurately, but it is still missing the lower part.

PBCGB takes on average 5.4 steps to go back to the home posture and 6.5 steps toward an end goal; it has 1937 end goals for 15000 iterations and goes back to the home posture 211 times. The variant with smaller discretization and higher maximum sub-goal distance, PBCGB-incgs, takes on average 4.8 steps to go back to the home posture and 2.8 steps toward an end goal; it has 4504 end goals for 15000 iterations, and goes back to the home posture 452 times.
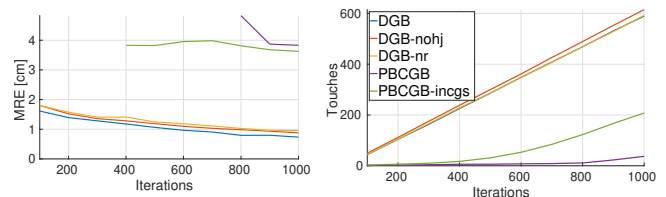

Fig. 15: Right hand reaching for the head – comparison of exploration strategies at 1000 iterations. (Left) MRE. (Right) Number of touches.

The mean percentage of discarded motor commands is given in Table. II.

DGB	DGB-nr	DGB-nohj	PBCGB	PBCGB-incgs
16.9±6.1	17.4±6.7	14.4±5.4	2.1±3.92	0.7±0.9

TABLE II: Mean percentage of discarded motor commands for head reaching. Relative to 1000 iterations, except PBCGB variants based on 15000 iterations.

V. SUMMARY AND DISCUSSION OF EXPERIMENTS

The results presented in this work are to our knowledge the first attempt to compare the intrinsic motivation framework of the "SAGG-RIAC" type (e.g., see [14]) with path-based continuous goal babbling approaches [22]. Furthermore, we study a nonstandard scenario of reaching to own body as opposed to reaching in space. There are a number of complications (see Section III-C and Fig. 5) having mostly to do with the fact that feedback about the learning progress is not available at all times: if no contact on the skin is generated, no error is available and learning as well as "external evaluation" are
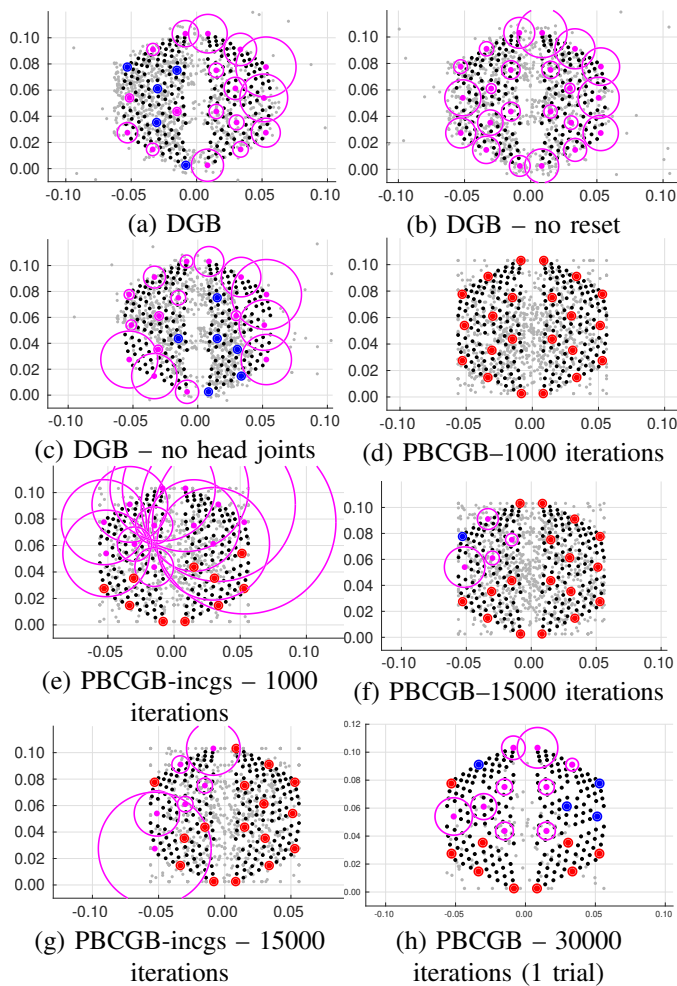
Fig. 16: Right hand reaching for the head – observation space. See Fig. 9 and Fig. 13 for details. (h) is only one trial that had been run to 30000 iterations (not a mean of 10 trials).

compromised. For this reason, we complemented the results presentation by reaching error (MRE) with the number of touches achieved and the goal space projection. None of these provide a complete picture, but together they improve our understanding. Additionally, the issue of discrepancy between target and actual joint configurations has to be tackled.

In line with our previous results [29], Discretized Goal Babbling (DGB) performed best. Motor space exploration methods—Random Motor Babbling and Motor Babbling with Intrinsic Motivation—performed worse than any Observation space exploration methods—Random or Discretized Goal Babbling. Beyond Explauto's DGB, the novel Path-Based Continuous Goal Babbling (PBCGB) is intended to learn a model that is both continuous and consistent. While these two key requirements are known to be met for reaching external objects [16], [22], the algorithm only achieves a partial exploration of the touch space. The key limitation of the current setup seems to be the discrete sensor layout with gaps between sensor patches and regions, which makes it hard for the path-based approach to discover the right local movement direction. The difficulty of the requirement for continuous motion is also demonstrated by the degeneration of DGB's performance when

generating continuous, "complex" touches in the "no reset" condition.

In summary, this work makes an important contribution in that it puts to test the algorithms developed for reaching in external space on a new problem: learning to reach for the agent's own skin surface. The fact that many reaching trials do not contact the skin and hence provide no training data for learning the inverse model poses a challenge. Additional practical complications come from two facts: (i) self-collisions can occur also between other body parts than the end effector and the target skin region; (ii) the skin sensor sheet is not continuous. The algorithms we tested proved to work in this situation but their performance was degraded. Sliding over the skin surface (or "complex touch") provides an opportunity for more efficient learning; however, due to the nature of our setup, it somewhat amplifies the effects of the practical complications mentioned above. A soft skin surface with overlapping tactile receptive fields would be a solution.

VI. GENERAL DISCUSSION AND FUTURE WORK

First, the issue of motor redundancy should be discussed. The inverse model—from skin space to joint space—was learned directly from the training samples and when using the Explauto methods, it was represented with the nearest neighbor algorithm. While *direct inverse modeling* [52] is prone to the ill-posedness of the general inverse kinematics problem and the averaging over non-convex solutions sets, our solution circumvents this by performing the exploration in the goal space: alternative solutions exploiting motor redundancy are thus not sought. Additionally, in the nearest neighbor algorithm, no averaging takes place. However, the solution found will in our case be the first solution found; it may thus depend on initialization or chance and may not be the best solution. *Distal learning*, or *learning with a distal teacher* [52], as opposed to direct inverse modeling, is more versatile in that it allows the incorporation of additional constraints to channel the search for the (single) solution. However, while initially a single solution to a reaching target on the body may suffice, we know that adults are capable of alternative solutions depending on context. Distal learning allows the incorporation of a forward model and inverse model in series. Such a solution is more versatile in that the forward model, which is unique, "disambiguates" between alternatives coming from the, one-to-many, inverse mappings and can check their correctness. Human motor control in the cerebellum may be employing multiple paired forward and inverse models [9] (see the MOSAIC model [53]). Distal learning can thus in principle deal with a redundant system, but the problem is that the motor error is not directly observable [16]. A solution that would allow the agent to find one solution for every reaching target first, but add and keep alternatives later on, remains our future work. The mixed—composite forward-inverse models—can be a solution (see [54] for a survey). In particular, associative memory architectures have been proposed that allow to store several solutions and dynamically switch between them [55]. In this way, it may be possible to construct a *continuous* model of touch that can perform uninterrupted motion throughout

complex touches, only flipping to another solution branch if necessary.

Self-touch configurations are also more kinematically constrained than reaching in free space in front of the body and hence the effective motor redundancy is likely lower. This is even more the case for the experiments used here, in which only five DoF of the Nao arm were employed. Other relevant extension points in recent work concern the exploitation of symmetries in the sensorimotor apparatus [56] that can lead to a reduced need for exploration. While previous work developed competence progress based goal babbling [14] and path based continuous goal babbling [22] separately, and based on a rather different overall organization of behavior, recent studies have demonstrated their compatibility [24]. The results of the present study regarding accuracy, but also consistency and continuity, suggest that also the (self-)touch domain could benefit from an integration of both approaches that prioritizes some paths over others.

Second, the use of the nearest neighbor algorithm for the inverse model representation has to be discussed. It has the following advantages: (i) incremental learning is simple and requires registering pairs from input and goal space only ("lazy learning"); (ii) there is no averaging or interpolation of samples (avoiding the problem of non-convexity of the solution space). The disadvantages are: (i) computational complexity: all experience is stored in memory and upon retrieval—query to the inverse model—time is required to find the nearest neighbor; (ii) susceptibility to noise: in our scenario, "phantom" skin activation would be catalogued together with the current joint configuration and contaminate the model. (iii) the mapping will not be smooth: adjoining skin receptors will not necessarily map to nearby joint configurations. Baranes and Oudeyer [14] deem nonparametric methods (like nearest neighbor) suitable for their problems (including inverse kinematics) and the complexity problem can be mitigated by efficient implementation [57]. Alternative representations of the inverse model could be local regression methods (e.g, Locally Weighted Projection Regression; Sigaud et al. [58] for a survey). How such mappings are encoded by the brain is an open question.

Third, the representations of the input and output spaces importantly influence what can be learned and how. In the input or motor space in this work, the actual execution of the movement—initiation, termination, and its dynamics—has not been addressed and such separation of movement preparation and control may not be justified [59]. Mannella et al. [27] do consider this aspect and observe, for example, that easy postures are acquired before hard ones. The computational perspective of Cisek could be used as a starting point to add this important dimension [60]. Dynamic Movement Primitives [61] could be employed and possibly allow the incorporation of rhythmic movements; Central Pattern Generators used in [40] are an alternative. Regarding the "skin space", one could come closer to the biological reality by mimicking the non-uniform density of receptors (as done in [25], [27]). On the representation level, self-organizing maps seem like a natural candidate [27], [62]. The distance metric required for the exploration will then be distorted as is typical for homuncular

representations and present an additional challenge. Furthermore, the spaces of torso and head were treated separately in this work. To make them part of the same goal space, a metric connecting them would have to be introduced. Alternatively, they could be separate goals, but within the same exploration framework. Forestier et al. [63] or Santucci et al. [64] provide possible solutions. Adding the skin space of the effector would bring additional complexity. Finally, the motor and sensory spaces could be treated in a more integrated manner as proposed by Marcel et al. [65] who present a mathematical analysis of building a sensorimotor representation of a naive agent's tactile space.

Fourth, it is also worth considering how the task studied here—reaching to the body—differs from reaching in general. As introduced in Section I-A, while learning to touch the body, feedback is not available for most arm postures as the arms are simply in free space. It is also important to consider the role this feedback plays. When the intrinsic motivation frameworks (e.g., see [14], [16]) are applied to reaching in general, an inverse model is being learned and the reaching error is used to monitor learning progress and channel it in different directions. An alternative formulation is offered by Caligiore et al. [66], employing reinforcement learning—the agent is rewarded for contacting an object in front of it. In addition, the reward is shaped such that the agent is motivated to reach the final position with a minimum speed, thus mimicking the minimum variance theory [67]. The third component of this model is the equilibrium point hypothesis (e.g., see [68]). Thus, this model produces reaching trajectories, while remaining at the level of reaching kinematics (circumventing forces and torques). Remarkably, a number of kinematic variables characterizing the development of reaching in infants is reproduced. However, it is not clear how the reinforcement learning component could be transferred to our case—if reward was delivered for any contact on the skin, the agent would not be motivated to explore the whole skin surface. Related to this is also the notion of goals in general. While in reinforcement learning frameworks the agent is maximizing a sum of future rewards, active goal exploration approaches like the one used here assume that the agent can "imagine" goals—positions on the skin here—and then attempt to achieve them. It remains an open question whether activation of tactile receptors could be interesting enough goals for the infant. Furthermore, and this applies to infant development in general, there is evidence suggesting that infants younger than two years cannot make certain kinds of goal-directed action—they can produce an action directed at a stimulus they see (using an inverse model), but they cannot "imagine" a stimulus and practice the action [69], [70]. A detailed discussion along with a computational model is presented in [8]. If this counts as evidence for our problem and given that we know the body exploration is mainly occurring during the first year of life, this would exclude the strategies that performed best—goal babbling with intrinsic motivation. However, this is still an open question and it cannot be excluded that infants' use of internal goals may be task- or context- dependent.

It is our ultimate goal to ground the model in biological data. The work of Schlesinger [71], investigating the looking

patterns of infants, is an example of such work. In our scenario, there are two concrete ways how we plan to proceed. First, in our study, the robot is learning an inverse model: which motor commands to use to reach to targets on its body. The performance for different body parts and at different stages of development can be compared with behavioral data from infants reaching for vibrotactile stimuli on the body [44]. For example, we should analyze how infants deal with the redundancy of their motor system in this particular case: during different "stages" in their development, do they use the same or distinct configurations to reach for targets on the body? If the latter were the case, the goal exploration strategies that suppress the redundancy of the motor system may not be appropriate. Also, with different initial postures, do infants tend to go to a canonical posture first? There is evidence suggesting that this may be the case in infants [72] and adults learning a new task [73]. A pilot analysis in this direction—following one infant from 4 to 18 months—has been presented here, suggesting that the reaching configurations are quite stereotypical. More quantitative analysis is needed.

Second, statistics obtained from studies observing spontaneous touches to the body in infants [35], [36]—such as how often infants touch particular body parts, in which sequence, etc.—could be fed into the robot simulator to train the inverse model and the results in terms of reaching performance to targets on the body compared with those obtained from the computational exploration strategies. Alternatively, we could aim to model the exploration process itself and obtain similar self-touch statistics as an emergent property. Discovering signatures of curiosity-driven learning in the brain is an active research area [74], employing fMRI [75] or EEG and body states [76]. Only behavioral data poses a greater challenge. With carefully designed experiments, one may be able to discern which cost function the "learning machine" is using [77]. Discriminating spontaneous vs. systematic exploration in naturalistic observations (like [35], [36]) remains to our knowledge an open question.

ACKNOWLEDGMENT

REFERENCES

[1] A. Yamaguchi and C. G. Atkeson, "Recent progress in tactile sensing and sensors for robotic manipulation: can we turn tactile sensing into vision?" *Advanced Robotics*, vol. 33, no. 14, pp. 661–673, 2019.

[2] C. Bartolozzi, L. Natale, F. Nori, and G. Metta, "Robots with a sense of touch," *Nature materials*, vol. 15, no. 9, pp. 921–925, 2016.

[3] S. Luo, J. Bimbo, R. Dahiya, and H. Liu, "Robotic tactile perception of object properties: A review," *Mechatronics*, vol. 48, pp. 54–67, 2017.

[4] G. Metta, L. Natale, F. Nori, G. Sandini, D. Vernon, L. Fadiga, C. von Hofsten, K. Rosander, M. Lopes, J. Santos-Victor, A. Bernardino, and L. Montesano, "The iCub humanoid robot: An open-systems platform for research in cognitive development," *Neural Networks*, vol. 23, no. 8-9, pp. 1125–1134, 2010.

[5] R. M. Bradley and C. M. Mistretta, "Fetal sensory receptors," *Physiological Reviews*, vol. 55, no. 3, pp. 352–382, 1975.

[6] A. Bremner and C. Spence, "The development of tactile perception," in *Advances in child development and behavior*. Elsevier, 2017, vol. 52, pp. 227–268.

[7] M. Hoffmann, "The role of self-touch experience in the formation of the self," in *The Development of the Self Workshop at IEEE ICDL-EpiRob*, 2017.

[8] L. Jacquey, G. Baldassarre, V. G. Santucci, and J. K. O'Regan, "Sensorimotor contingencies as a key drive of development: from babies to robots," *Frontiers in neurorobotics*, vol. 13, p. 98, 2019.

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

[10] A. N. Meltzoff and M. K. Moore, "Explaining facial imitation: A theoretical model," *Infant and child development*, vol. 6, no. 3-4, pp. 179–192, 1997.

[11] D. Bullock, S. Grossberg, and F. H. Guenther, "A self-organizing neural model of motor equivalent reaching and tool use by a multijoint arm," *Journal of Cognitive Neuroscience*, vol. 5, no. 4, pp. 408–435, 1993.

[12] J. Schmidhuber, "A possibility for implementing curiosity and boredom in model-building neural controllers," in *Proc. of the International Conference on Simulation of Adaptive Behavior: From Animals to Animats*, 1991, pp. 222–227.

[13] P.-Y. Oudeyer and F. Kaplan, "What is intrinsic motivation? A typology of computational approaches," *Frontiers in neurorobotics*, vol. 1, p. 6, 2007.

[14] A. Baranes and P.-Y. Oudeyer, "Active learning of inverse models with intrinsically motivated goal exploration in robots," *Robotics and Autonomous Systems*, vol. 61, no. 1, pp. 49–73, 2013.

[15] G. Baldassarre and M. Mirolli, *Intrinsically motivated learning in natural and artificial systems*. Springer, 2013.

[16] M. Rolf, J. J. Steil, and M. Gienger, "Goal babbling permits direct learning of inverse kinematics," *IEEE Trans. Auton. Mental Develop.*, vol. 2, no. 3, pp. 216–229, 2010.

[17] V. Y. Roschin, A. A. Frolov, Y. Burnod, and M. A. Maier, "A neural network model for the acquisition of a spatial body scheme through sensorimotor interaction," *Neural computation*, vol. 23, no. 7, pp. 1821–1834, 2011.

[18] M. Rolf, "Goal babbling for an efficient bootstrapping of inverse models in high dimensions," Ph.D. dissertation, Bielefeld University, 2012.

[19] R. M. Ryan and E. L. Deci, "Intrinsic and extrinsic motivations: Classic definitions and new directions," *Contemporary educational psychology*, vol. 25, no. 1, pp. 54–67, 2000.

[20] A. Baranes and P.-Y. Oudeyer, "R-iac: Robust intrinsically motivated exploration and active learning," *IEEE Transactions on Autonomous Mental Development*, vol. 1, no. 3, pp. 155–169, 2009.

[21] K. Narioka and J. J. Steil, "U-shaped motor development emerges from goal babbling with intrinsic motor noise," in *2015 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. IEEE, 2015, pp. 55–62.

[22] M. Rolf and J. J. Steil, "Efficient exploratory learning of inverse kinematics on a bionic elephant trunk," *IEEE transactions on neural networks and learning systems*, vol. 25, no. 6, pp. 1147–1160, 2013.

[23] C. Moulin-Frier, P. Rouanet, and P.-Y. Oudeyer, "Explauto: an open-source python library to study autonomous exploration in developmental robotics," in *4th International Conference on Development and Learning and on Epigenetic Robotics*. IEEE, 2014, pp. 171–172.

[24] R. Rayyes, H. Donat, and J. Steil, "Efficient online interest-driven exploration for developmental robots," *IEEE Transactions on Cognitive and Developmental Systems*, 2020.

[25] H. Mori and Y. Kuniyoshi, "A human fetus development simulation: Self-organization of behaviors through tactile sensation," in *Development and Learning (ICDL), 2010 IEEE 9th International Conference on*. IEEE, 2010, pp. 82–87.

[26] Y. Yamada, H. Kanazawa, S. Iwasaki, Y. Tsukahara, O. Iwata, S. Yamada, and Y. Kuniyoshi, "An embodied brain model of the human foetus," *Scientific Reports*, vol. 6, 2016.

[27] F. Mannella, V. G. Santucci, E. Somogyi, L. Jacquey, K. J. O'Regan, and G. Baldassarre, "Know your body through intrinsic goals," *Frontiers in Neurorobotics*, vol. 12, p. 30, 2018.

[28] D. Caligiore and G. Baldassarre, "The development of reaching and grasping: Towards an integrated framework based on a critical review of computational and robotic models," in *Reach-to-grasp Behavior: Brain, Behavior, and Modelling Across the Life Span*, D. Corbetta and M. Santello, Eds. Routledge, 2018, pp. 335–364.

[29] F. Gama, M. Shcherban, M. Rolf, and M. Hoffmann, "Active exploration for body model learning through self-touch on a humanoid robot with artificial skin," in *Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 2020.

[30] A. Piontelli, *Development of Normal Fetal Movements; The Last 15 Weeks of Gestation*. Italy: Springer-Verlag, 2015.

[31] M. Myowa-Yamakoshi and H. Takeshita, "Do human fetuses anticipate self-oriented actions? a study by four-dimensional (4d) ultrasonography," *Infancy*, vol. 10, no. 3, pp. 289–301, 2006.

[32] S. Zoia, L. Blason, G. D'Ottavio, M. Bulgheroni, E. Pezzetta, A. Scabar, and U. Castiello, "Evidence of early development of action planning in the human foetus: a kinematic study," *Experimental Brain Research*, vol. 176, no. 2, pp. 217–226, 2007.

[33] P. Rochat, "Hand-mouth coordination in the newborn: Morphology, determinants, and early development of a basic act," *Advances in psychology*, vol. 97, pp. 265–288, 1993.

[34] ——, "Self-perception and action in infancy," *Experimental brain research*, vol. 123, no. 1-2, pp. 102–109, 1998.

[35] B. L. Thomas, J. M. Karl, and I. Q. Whishaw, "Independent development of the reach and the grasp in spontaneous self-touching by human infants in the first 6 months," *Frontiers in psychology*, vol. 5, p. 1526, 2015.

[36] A. DiMercurio, J. P. Connell, M. Clark, and D. Corbetta, "A naturalistic observation of spontaneous touches to the body and environment in the first 2 months of life," *Frontiers in psychology*, vol. 9, p. 2613, 2018.

[37] G. Z. Tau and B. S. Peterson, "Normal development of brain circuits," *Neuropsychopharmacology*, vol. 35, no. 1, p. 147, 2010.

[38] J. Piaget, *The origins of intelligence in children*. International University Press New York, 1952.

[39] A. Van der Meer, F. Van der Weel, and D. N. Lee, "The functional significance of arm movements in neonates," *Science*, vol. 267, no. 5198, pp. 693–695, 1995.

[40] V. C. Meola, D. Caligiore, V. Sperati, L. Zollo, A. L. Ciancio, F. Taffoni, E. Guglielmelli, and G. Baldassarre, "Interplay of rhythmic and discrete manipulation movements during development: A policy-search reinforcement-learning robot model," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 8, no. 3, pp. 152–170, 2016.

[41] L. K. Chinn, M. Hoffmann, J. E. Leed, and J. J. Lockman, "Reaching with one arm to the other: Coordinating touch, proprioception, and action during infancy," *Journal of Experimental Child Psychology*, vol. 183, pp. 19–32, 2019.

[42] L. K. Chinn, C. F. Noonan, and J. J. & Lockman, "The human face becomes mapped as a sensorimotor reaching space during the first year," *Child Development*, 2020.

[43] M. Hoffmann, L. K. Chinn, E. Somogyi, T. Heed, J. Fagard, J. J. Lockman, and J. K. O'Regan, "Development of reaching to the body in early infancy: From experiments to robotic models," in *Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 2017, pp. 112–119.

[44] J. E. Leed, L. K. Chinn, and J. J. Lockman, "Reaching to the self: The development of infants' ability to localize targets on the body," *Psychological science*, vol. 30, no. 7, pp. 1063–1073, 2019.

[45] E. Somogyi, L. Jacquey, T. Heed, M. Hoffmann, J. Lockman, L. Granjon, J. Fagard, and J. K. O'Regan, "Which limb is it? responses to vibrotactile stimulation in early infancy," *British Journal of Developmental Psychology*, vol. 36, no. 3, pp. 384–401, 2018.

[46] J. Kelso, J. Buchanan, and T. Murata, "Multifunctionality and switching in the coordination dynamics of reaching and grasping," *Human Movement Science*, vol. 13, no. 1, pp. 63–94, 1994.

[47] P. Maiolino, M. Maggiali, G. Cannata, G. Metta, and L. Natale, "A flexible and robust large scale capacitive tactile system for robots," *Sensors Journal, IEEE*, vol. 13, no. 10, pp. 3910–3917, 2013.

[48] M. Shcherban, 2020 nao-gazebo-skin. [Online]. Available: https://github.com/maxymczech/nao-gazebo-skin

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

[50] C. Moulin-Frier and P.-Y. Oudeyer, "Exploration strategies in developmental robotics: a unified probabilistic framework," in *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)*. IEEE, 2013, pp. 1–6.

[51] P. Loviken and N. Hemion, "Online-learning and planning in high dimensions with finite element goal babbling," in *2017 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. IEEE, 2017, pp. 247–254.

[52] M. I. Jordan and D. E. Rumelhart, "Forward models: Supervised learning with a distal teacher," *Cognitive science*, vol. 16, no. 3, pp. 307–354, 1992.

[53] M. Haruno, D. M. Wolpert, and M. Kawato, "Mosaic model for sensorimotor learning and control," *Neural computation*, vol. 13, no. 10, pp. 2201–2220, 2001.

[54] D. Nguyen-Tuong and J. Peters, "Model learning for robot control: a survey," *Cognitive processing*, vol. 12, no. 4, pp. 319–340, 2011.

[55] R. F. Reinhart and M. Rolf, "Learning versatile sensorimotor coordination with goal babbling and neural associative dynamics," in *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)*. IEEE, 2013, pp. 1–7.

[56] R. Rayyes, D. Kubus, and J. Steil, "Learning inverse statics models efficiently with symmetry-based exploration," *Frontiers in neurorobotics*, vol. 12, p. 68, 2018.

[57] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration." *VISAPP (1)*, vol. 2, no. 331-340, p. 2, 2009.

[58] O. Sigaud, C. Salaun, and V. Padois, "On-line regression algorithms for learning mechanical models of robots: A survey," *Robotics and Autonomous Systems*, vol. 59, no. 12, pp. 1115 – 1129, 2011.

[59] G. Schöner, J. Tekülve, and S. Zibner, "Reaching for objects: a neural process account in a developmental perspective," in *Reach-to-grasp Behavior: Brain, Behavior, and Modelling Across the Life Span*, D. Corbetta and M. Santello, Eds. Routledge, 2018.

[60] P. E. Cisek, "A computational perspective on proprioception and movement guidance in parietal cortex," in *The Somatosensory System: Deciphering the Brain's Own Body Image*. CRC Press, 2001, pp. 275–297.

[61] A. J. Ijspeert, J. Nakanishi, and S. Schaal, "Learning attractor landscapes for learning motor primitives," in *Advances in neural information processing systems*, 2003, pp. 1547–1554.

[62] M. Hoffmann, Z. Straka, I. Farkas, M. Vavrecka, and G. Metta, "Robotic homunculus: Learning of artificial skin representation in a humanoid robot motivated by primary somatosensory cortex," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 2, pp. 163–176, 2018.

[63] S. Forestier, R. Portelas, Y. Mollard, and P.-Y. Oudeyer, "Intrinsically motivated goal exploration processes with automatic curriculum learning," *arXiv preprint arXiv:1708.02190*, 2017.

[64] V. G. Santucci, G. Baldassarre, and M. Mirolli, "Grail: a goal-discovering robotic architecture for intrinsically-motivated learning," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 8, no. 3, pp. 214–231, 2016.

[65] V. Marcel, S. Argentieri, and B. Gas, "Building a sensorimotor representation of a naive agent's tactile space," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 9, no. 2, pp. 141–152, 2016.

[66] D. Caligiore, D. Parisi, and G. Baldassarre, "Integrating reinforcement learning, equilibrium points, and minimum variance to understand the development of reaching: A computational model." *Psychological review*, vol. 121, no. 3, p. 389, 2014.

[67] C. M. Harris and D. M. Wolpert, "Signal-dependent noise determines motor planning," *Nature*, vol. 394, no. 6695, pp. 780–784, 1998.

[68] A. G. Feldman, "Once more on the equilibrium-point hypothesis (λ model) for motor control," *Journal of motor behavior*, vol. 18, no. 1, pp. 17–54, 1986.

[69] U. Klossek, J. Russell, and A. Dickinson, "The control of instrumental action following outcome devaluation in young children aged between 1 and 4 years." *Journal of Experimental Psychology: General*, vol. 137, no. 1, p. 39, 2008.

[70] B. Kenward, S. Folke, J. Holmberg, A. Johansson, and G. Gredebäck, "Goal directedness and decision making in infants." *Developmental psychology*, vol. 45, no. 3, p. 809, 2009.

[71] M. Schlesinger, "Investigating the origins of intrinsic motivation in human infants," in *Intrinsically motivated learning in natural and artificial systems*. Springer, 2013, pp. 367–392.

[72] N. E. Berthier, R. K. Clifton, D. D. McCall, and D. J. Robin, "Proximodistal structure of early reaching in human infants," *Experimental Brain Research*, vol. 127, no. 3, pp. 259–269, 1999.

[73] M. Rohde, K. Narioka, J. J. Steil, L. K. Klein, and M. O. Ernst, "Goal-related feedback guides motor exploration and redundancy resolution in human motor skill acquisition," *PLoS Computational Biology*, vol. 15, no. 3, p. e1006676, 2019.

[74] J. Gottlieb and P.-Y. Oudeyer, "Towards a neuroscience of active sampling and curiosity," *Nature Reviews Neuroscience*, vol. 19, no. 12, pp. 758–770, 2018.

[75] M. J. Gruber, B. D. Gelman, and C. Ranganath, "States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit," *Neuron*, vol. 84, no. 2, pp. 486–496, 2014.

[76] A. Appriou, J. Ceha, E. Law, P.-Y. Oudeyer, and F. Lotte, "Towards measuring states of curiosity through electroencephalography and body sensors responses," in *CORTICO*, 2019.

[77] J. G. Cashaback, H. R. McGregor, A. Mohatarem, and P. L. Gribble, "Dissociating error-based and reinforcement-based loss functions during sensorimotor learning," *PLoS Computational Biology*, vol. 13, no. 7, p. e1005623, 2017.