Learning the Body Schema: A Developmentally-Inspired Framework for Curiosity-Driven Multimodal Self-Exploration

1st Lucas Yang
Department of Computer Science
University College London
London, United Kingdom
lucas.yang.24@ucl.ac.uk

***Abstract*—Human infants autonomously learn a sophisticated internal model of their own bodies, or a 'body schema', through spontaneous, intrinsically motivated self-exploration. Replicating this remarkable developmental process in robotics presents a grand challenge, moving beyond pre-specified models towards true autonomous learning. This paper addresses this challenge by proposing a novel, developmentally-inspired framework that enables a robot to learn its body schema in a goal-free, open-ended manner. Our approach integrates three key components: (1) a probabilistic body schema constructed using a Variational Autoencoder (VAE) to fuse proprioceptive and tactile information; (2) an Intrinsic Curiosity Module (ICM) that generates exploration rewards from prediction errors in the learned latent space; and (3) a novel dynamic touch reward mechanism that simulates infant habituation and interest shift to guide exploration. Without any prior knowledge of its body structure, our simulated infant robot develops structured exploratory behaviors from random motor babbling, including an emergent rostro-caudal progression that is remarkably consistent with patterns observed in human infants. Our work provides a concrete, computable model for the underlying mechanisms of early body cognition and presents a viable pathway for creating robots that learn and develop autonomously.**

***Index Terms*—Developmental Robotics, Intrinsic Motivation, Body Schema, Self-Exploration, Reinforcement Learning, Curiosity-Driven Learning, Multimodal Learning**

I. Introduction

The process of human cognitive development is one of nature's most remarkable phenomena. An infant, without any explicit task guidance, driven solely by innate curiosity and an urge to explore, progressively masters complex motor skills and builds a profound understanding of itself and the world. This capacity for autonomous learning and development stands in stark contrast to the current state of artificial intelligence and robotics. Conventional robotic systems typically rely on pre-programmed instructions or models meticulously engineered for specific tasks; they lack the ability to learn and develop in an open-ended manner from the ground up [1].

Therefore, a core ambition within robotics and artificial intelligence is to create agents capable of **autonomous learning and development**, mirroring the process seen in human infants. To achieve this grand goal, an inescapable and fundamental starting point is for the agent to first understand itself—by constructing its own **Body Schema** [2], [3]. The body schema is not a static geometrical model, but rather a dynamic, action-oriented sensorimotor representation system [4]. Developmental psychology research provides compelling evidence that this complex internal model is not innate, but is progressively learned from the high-dimensional, multimodal stream of sensory information generated through extensive spontaneous movements and self-touch [5], [6].

Inspired by this, this paper aims to answer a core scientific question: **Can a robot, in the absence of any external goals, autonomously learn and form an effective body schema solely through intrinsically motivated, multimodal self-exploration?**

Addressing this question requires overcoming two primary challenges. First is the problem of motivation for exploration. In unsupervised scenarios lacking external tasks or explicit rewards, what is the fundamental driving force for an agent to perform exploratory actions? Many paradigms, such as **Goal Babbling** [7], [8], rely on competence-based motivation within a predefined goal space. However, this goal-oriented framework differs substantially from the apparently goal-free, open-ended exploration seen in early infancy, which appears to be driven by knowledge-based motivations such as curiosity [9]. Second is the challenge of multimodal information representation. The robot must learn a compact, effective, and predictive body representation from high-dimensional proprioceptive and extremely sparse tactile sensory data [10].

Inspired by the developmental process in infants, this paper proposes a novel, end-to-end autonomous learning framework to address these challenges. Our core idea is not to provide the robot with any external goals, but to drive its learning and exploration through an internal, prediction-based cognitive loop. Our framework is built upon three key technical pillars:

1) We employ a **Variational Autoencoder (VAE)** to construct a probabilistic, generative body schema [11]. This model fuses and compresses multimodal sensory inputs into a low-dimensional latent state space [12], upon which a predictive **Forward Model** is learned [13].
2) Exploration is driven by an **Intrinsic Curiosity Module (ICM)**, where the intrinsic reward is derived directly

from the prediction error of the body schema's forward model [14]. This mechanism is a computationally tractable implementation of the cognitive principle of rewarding "learning progress" [9], [15].

3) To effectively focus exploration on the body itself, we have designed a novel **dynamic touch encouragement mechanism**. This mechanism mathematically simulates the phenomena of **habituation** and **interest shift** observed in developmental psychology [6], [16], guiding the robot to continually explore new areas of its body.

Through the interplay of these simple, local learning rules, our framework demonstrates the autonomous emergence of globally structured, developmentally plausible exploratory behaviors, including a rostro-caudal progression that is remarkably consistent with patterns observed in human infants.

The main contributions of this paper are as follows:

1) We propose a unified, intrinsically-driven learning framework that enables a robot to autonomously learn its body schema in a goal-free, open-ended manner.
2) We design a novel, dynamic touch reward mechanism inspired by biological developmental principles, which effectively guides the diversity and completeness of the exploration.
3) We provide a quantitative analysis of the emergent transition from random movements to structured, infant-like self-touch behaviors.
4) We present experimental evidence for a developmental shift in the robot's learning drivers, validating the efficacy of the proposed curriculum.

The remainder of this paper is organized as follows: Section II reviews related work. Section III details the proposed methodology. Section IV presents the experimental setup and results, followed by discussion and conclusion in Sections V and VI.

II. Related Work

A. The Developmental Blueprint: Infant Self-Exploration as a Benchmark

Our work is deeply inspired by the cognitive development processes in human infants. In the absence of explicit task guidance, infants progressively construct a perception of their own bodies through spontaneous exploratory behaviors, providing a valuable "developmental blueprint" for robotics [1]. The first few months of life are critical to this process, during which infants exhibit high-frequency, seemingly random spontaneous movements, known as "Motor Babbling," and extensively contact their own bodies through these actions [4].

Recent longitudinal studies have provided rich quantitative evidence for these early behaviors. For instance, research by DiMercurio et al. [6] found that infants as young as 3 to 9 weeks old can produce up to 200 contacts with their body and supporting surfaces within a 10-minute observation period. These tactile interactions are not entirely random but follow specific developmental patterns. A prominent trend is the "rostro-caudal" progression of exploration: in the earliest weeks, touches are concentrated on the head and torso; as infants age, the target of touch gradually moves toward the lower body, with a significant increase in contacts to the legs and feet around 14 weeks of age [5].

From a cognitive science perspective, these spontaneous explorations are a core mechanism for the construction of the infant's body schema. It is theorized that self-touch provides a unique "double touch" sensation (simultaneously feeling the touch and being touched) [4]. Furthermore, a fundamental learning mechanism observable from birth is habituation: infants' attention to a repeated stimulus gradually decreases, while their attention is recaptured by novel stimuli [16], [17]. This innate drive for novelty provides a psychological foundation for our knowledge-based intrinsic motivation framework.

In summary, these findings from infant studies provide the core design specifications for our computational model: the learning process must be spontaneous and unsupervised, and the emergent exploratory behavior should exhibit a progressive, developmentally-plausible character.

B. Computational Approaches to Autonomous Body Schema Learning

Inspired by this developmental blueprint, robotics research has explored various computational paradigms for autonomous body schema acquisition. These approaches differ primarily in how they represent the body and what form of intrinsic motivation drives the exploration.

Body Schema Representation. Approaches to modeling the body schema can be broadly categorized as explicit, implicit, and generative. Explicit models rely on a predefined structure, such as a kinematic chain, and focus on online parameter calibration, which offers adaptability but is constrained by the initial model design [18]. Implicit models learn a direct sensor-to-motor mapping without structural assumptions, but are often data-hungry and lack interpretability [19]. Our work aligns with the more recent and promising direction of **generative representation** [20]. By using a Variational Autoencoder (VAE) [11], we learn the underlying low-dimensional probability distribution of the entire sensorimotor data stream. This allows for effective fusion of multimodal information and the construction of a predictive internal model [12], innovatively directing this powerful technique towards learning the robot's 'internal world'—its own body.

Intrinsic Motivation Paradigms. A critical distinction in exploration strategies lies between goal-directed and goal-free paradigms. **Goal-directed (or competence-based) motivation** has proven highly effective. The Goal Babbling framework, for instance, dramatically enhances learning efficiency by exploring directly in a task-relevant goal space rather than the high-dimensional motor space [7]. Subsequent work has enabled robots to autonomously select goals that maximize "learning progress," creating a self-organized curriculum [8], [21]. While powerful, the fundamental limitation of these methods is their reliance on a goal-oriented framework, where goals are either predefined or segmented from the sensory

space. This paradigm contrasts sharply with the open-ended, seemingly goal-free nature of early infant exploration.

Our work is founded on the alternative paradigm of **knowledge-based motivation**, which does not depend on explicit goals. This approach encourages an agent to acquire new knowledge about its world model, driven by principles like curiosity [9]. The theoretical basis for this idea is that agents are intrinsically rewarded by the "compression progress" of their internal models [15]. A powerful and elegant implementation of this concept is the **Intrinsic Curiosity Module (ICM)**, proposed by Pathak et al. [14]. The ICM defines curiosity as the error in an agent's ability to predict the consequences of its own actions. It consists of a *forward dynamics model*, which predicts the next state in a feature space, and its prediction error serves as the intrinsic reward. It also includes an *inverse dynamics model*, which learns this feature space by predicting the action that caused the state transition, a process that cleverly makes the features robust to uncontrollable "white noise" in the environment. We argue that this knowledge-based, goal-free paradigm is more developmentally plausible for modeling the foundational stages of body schema acquisition. Our framework builds upon this principle, uniquely combining a curiosity drive with a developmentally-inspired touch encouragement mechanism to guide the open-ended exploration process.

III. METHODOLOGY

In this section, we detail a developmentally-inspired, end-to-end reinforcement learning framework for autonomous body schema acquisition. The core of this framework is an intrinsically motivated agent that learns a probabilistic body model and develops structured exploratory behaviors through interaction with its own body.

A. Overall Learning Framework

Our approach is built upon the framework of deep reinforcement learning. The agent's behavioral policy, π, is optimized using the Proximal Policy Optimization (PPO) algorithm [22]. Unlike traditional reinforcement learning, our agent does not rely on any external task rewards; its behavior is driven entirely by a synergistic, hybrid intrinsic reward signal that we have designed.

As illustrated in Fig. 1, the entire system operates as a closed-loop interactive process. At each timestep t, the agent's policy π generates an action a_t based on the multimodal state representation z_t from the Body Schema Model. This action is executed in the environment, yielding a new raw observation o_{t+1} (comprising proprioception and touch). This new observation, along with the previous one, is used to update the Body Schema Model. Concurrently, the Synergistic Intrinsic Reward Module computes a total intrinsic reward R_t. This reward signal is ultimately fed back to the PPO algorithm to update the policy, thereby guiding the next round of behavior.

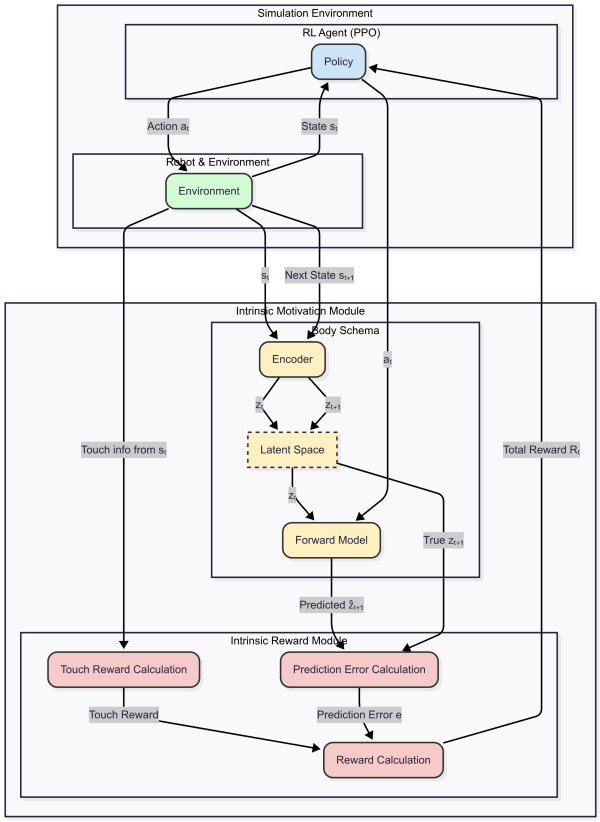


Fig. 1: The overall architecture of our proposed autonomous learning framework, illustrating the information flow between the agent, the environment, the Body Schema Model, and the Synergistic Intrinsic Reward Module.

B. Probabilistic Body Schema Model

The core of the agent is a probabilistic, generative body schema model, which is responsible for fusing and compressing high-dimensional multimodal sensory inputs and learning the body's internal dynamics.

1) Dual-Channel VAE Architecture for State Representation: Given the heterogeneity of the proprioceptive (dense) and tactile (sparse) data, we employ a dual-channel Variational Autoencoder (VAE) architecture to process them separately [11]. A $VAE_{proprio}$ and a VAE_{touch} encode their respective raw inputs, $o_{proprio,t}$ and $o_{touch,t}$. The means of their respective latent distributions are then concatenated to form a unified, low-dimensional latent body state representation z_t. This vector z_t constitutes the core state space of our body schema. The VAEs are trained to optimize the Evidence Lower Bound (ELBO), which consists of a reconstruction loss term to ensure information quality and a KL-divergence regularization term to structure the latent space.

2) Forward Dynamics Model in Latent Space: We learn a Forward Model f_θ in the latent space, which is implemented as a Multi-Layer Perceptron (MLP). This model serves as an "emulator" of the body's dynamics as described by Grush [13] and forms the basis for generating intrinsic curiosity. The

model employs a residual architecture, meaning it predicts the *change* in state Δz_t rather than the next state itself: $\hat{z}_{t+1} = z_t + f_\theta(z_t, a_t)$. This structure is known to stabilize the learning process, particularly when state transitions are smooth.

3) Training with Stabilized Loss Function: The entire body schema model (including both VAEs and the Forward Model) is trained in batches at the end of each rollout using the collected experience data. To stabilize the training process, all our predictive models employ a specialized loss function derived from the negative log-likelihood of a Gaussian distribution. This loss, which we term σ-Loss, is proportional to the logarithm of the Mean Squared Error (MSE): $L_\sigma = \log(\text{MSE}(\text{prediction, target}) + \epsilon)$. This design prevents excessively large gradients when the prediction error is high and provides a smoother gradient landscape, thereby enhancing overall learning stability.

C. Synergistic Intrinsic Reward Mechanism

The agent's exploration is driven entirely by a carefully designed hybrid intrinsic reward system, which synergistically combines knowledge-based curiosity with developmentally-inspired touch encouragement.

1) Total Reward and Dynamic Weighting Curriculum: At each timestep t, the total intrinsic reward R_t is a weighted sum of three primary components:

$$R_t = \lambda_{icm}(t) \cdot R_{icm,t} + \lambda_{touch}(t) \cdot R_{touch,t} + \lambda_{hand}(t) \cdot R_{hand,t} \tag{1}$$

The weighting coefficients $\lambda(t)$ are not fixed but follow a curriculum learning [23] schedule. The weights are linearly interpolated over the initial phase of training until a predefined step count is reached, after which they remain constant. This curriculum guides the agent to first focus on the simpler task of "discovering" its body through touch (dominated by touch rewards), and then gradually shifts the focus to the more complex task of "understanding" its body's dynamics through curiosity (dominated by the ICM reward).

2) ICM Curiosity Reward ($R_{icm,t}$): This reward component stems from knowledge-based intrinsic motivation [9]. $R_{icm,t}$ is directly defined as the prediction error of the forward model, L_{fwd}, computed in the latent space. This reward incentivizes the agent to explore state-action regions where its internal body model is still inaccurate, effectively pushing it to explore its "cognitive frontier."

Crucially, unlike the original ICM formulation for visually complex environments [14], our framework does not require an inverse dynamics model. This is because our agent's observation space—proprioception and touch—is fully determined by its own motor commands. There are no uncontrollable external factors (e.g., unpredictable moving objects in a visual scene) that could act as "distractors." Therefore, the latent space learned by the VAE is already a sufficient representation of the agent's controllable state, allowing for a more direct and computationally efficient implementation of the curiosity module.

3) Dynamic Touch Encouragement Reward ($R_{touch,t}$ and $R_{hand,t}$): This component is designed to provide a focused, yet non-repetitive, incentive for self-touch, simulating key concepts and phenomena from developmental psychology.

First, to model the unique "double touch" sensory feedback experienced by infants during self-exploration [4] and to prevent the agent from falling into behavioral traps like "touching the floor," we distinguish between **body touches** (e.g., hand-to-torso) and **hand touches** (e.g., hand-to-hand). Both categories share a basic set of **temporal gating mechanisms**, including a brief "reward window" to validate contact, a "cooldown period" to prevent continuous triggering from a single sensor, and an "overhold penalty" to discourage static, non-exploratory contact. The specific parameters for these gates differ between the two touch types to encourage a variety of interactions.

Building on this foundation, we introduce a more sophisticated reward modulation mechanism for **body touches** to simulate the core dynamics of infant learning:

- **Habituation** [16], [17]: To prevent the agent from fixating on easily accessible body parts, we maintain a history of recent touches for each body region. The potential reward for a body part undergoes an **exponential decay** based on its recent touch frequency.
- **Interest Shift** [6]: To encourage comprehensive exploration, the decayed potential rewards for all body parts are normalized via a **Softmax function**. This process ensures that as one area becomes "boring" (its reward potential decays), the reward "budget" is automatically reallocated to other, less-explored, "novel" areas, guiding the agent to shift its exploratory focus.

In summary, this complex mechanism, combining temporal gating, exponential decay, and Softmax normalization, not only focuses the agent's exploration but, more importantly, computationally replicates the processes of habituation and interest shift that drive efficient and diverse exploration in human infants.

IV. Experiments and Results

In this section, we first detail the experimental setup, including the simulation platform, agent specifications, key training parameters, and the design of our ablation studies. Subsequently, we present a comprehensive analysis of the results, beginning with the remarkable emergent behaviors of our complete framework (the baseline model). Finally, through rigorous comparison with five ablation studies, we systematically dissect our framework to demonstrate the critical role and necessity of each core component.

A. Experimental Setup

1) Simulation Platform and Agent: Our experiments were conducted within the **Babybench** simulation platform, an open-source environment expressly designed for research in developmental robotics [24]. Babybench leverages the **MuJoCo** physics engine to provide high-fidelity physical interaction simulations [25]. The agent used is the platform's built-in infant robot model, **MIMo (Multi-Modal Infant Model)** [26].

Morphologically and dimensionally analogous to a human infant, MIMo features **30 degrees of freedom (DoF)**, providing a high-dimensional, continuous action space conducive to learning complex behaviors. To facilitate comparison with developmental timelines, we note that 1000 simulation steps in our setup correspond to approximately 16 seconds of real-world time.

2) State and Action Space: The agent's raw observation o_t is a multimodal input composed of two primary sensory modalities, configured as follows:

- **Proprioception**: A 264-dimensional vector ($o_{proprio} \in \mathbb{R}^{264}$), which is a concatenation of the robot's joint positions, angular velocities, sensed torques, joint limit states, and current actuation values. This provides a comprehensive snapshot of the robot's internal state.
- **Touch**: A 17169-dimensional, high-dimensional sparse vector ($o_{touch} \in \mathbb{R}^{17169}$). This signal represents the readings from the artificial skin taxels covering the entire body.

The agent's action space a_t is a 30-dimensional continuous vector, corresponding to target position commands for the controller of each of the 30 actuated joints.

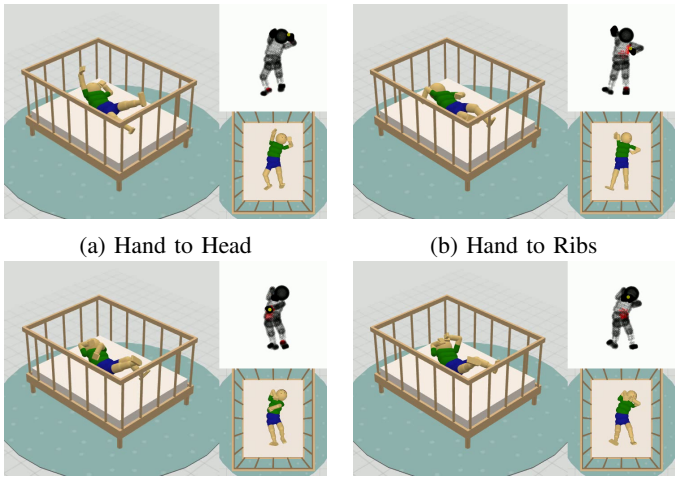
3) Training Parameters: We employ the **PPO (Proximal Policy Optimization)** algorithm from the Stable Baselines 3 library as our core reinforcement learning trainer [22]. The policy network is a `MultiInputPolicy`, configured to process proprioceptive and tactile modalities separately. Unless otherwise specified, all experiments were trained for **4 million timesteps**. Key hyperparameters are detailed in Table I.

4) Ablation Studies: To validate the efficacy of each component within our framework, we designed five ablation studies and compared their performance against the complete framework, hereafter referred to as the **Baseline** model:

1) **No Touch Reward**: The touch reward weights were persistently set to zero. The agent was driven solely by the ICM's curiosity reward.
2) **No Curiosity**: The ICM reward weight was set to zero. The agent was driven only by our dynamic touch reward mechanism.
3) **No Dynamic Weights**: The reward weights were fixed from the start to their final values in the baseline model's curriculum.
4) **Simple Touch Reward**: The dynamic touch reward mechanism was replaced with a simple, fixed reward for each tactile sensor activation.
5) **Simple MLP Model**: The VAE-based representation module was replaced with an end-to-end MLP.

B. Results and Analysis

1) Baseline Model: Emergence of Infant-like Exploratory Behavior: Our primary finding is that the **Baseline** model, operating without any predefined goals or external rewards, autonomously transitions from initial random motor babbling to structured, diverse, and developmentally plausible exploratory behaviors. As illustrated in Fig. 2, the agent discovers a rich repertoire of self-touch patterns, forming the basis for its body schema learning.



(a) Hand to Head (b) Hand to Ribs

(c) Cross-lateral Touch (d) Hand to Hand

Fig. 2: Examples of self-touch behaviors exhibited by the MIMo robot during exploration, including contact to the head, torso, cross-lateral limbs, and hands.

Most strikingly, the agent's exploration spontaneously organizes into an ordered sequence that closely mirrors the **rostro-caudal (head-to-tail) developmental gradient** observed in human infants. As shown in Fig. 3, during the initial 1 million steps (approximating the first 4.5 hours of experience), touch interactions are predominantly concentrated on the `head` and upper torso (`ub3`). Subsequently, the focus of exploration progressively shifts downwards, with a marked increase in contacts to the lower back/pelvis (`lb`) and waist (`cb`) in later stages. This emergent developmental trajectory is remarkably consistent with longitudinal studies of infants, which report a similar shift from head and torso exploration in the first few weeks to lower-body exploration around 14 weeks of age [6], [27].

The overall evolution of touch distribution, illustrated in Fig. 4, further confirms this trend, showing a clear diversification from an upper-body-centric pattern to a comprehensive, full-body exploration over the course of training.

This structured behavior is underpinned by two key learning dynamics. First, the internal body schema model demonstrates continuous and effective learning. Fig. 5 shows that both the VAE reconstruction losses and the forward model prediction loss consistently decrease, indicating the agent is successfully building a compact and predictive representation of its own body. Second, we observe a clear developmental crossover in the agent's motivational drivers, as depicted in Fig. 6. Initially, the touch rewards act as a crucial behavioral "anchor," guiding the agent to discover self-contact. As the agent masters this basic skill, the touch reward's influence wanes (due to habituation), while the curiosity-driven ICM reward becomes the dominant driver. This reveals a cognitive shift from a

TABLE I: Key Hyperparameters for the Training and Model Architecture

Component	Parameter	Value
PPO Algorithm Parameters		
	Learning Rate	1×10^{-4}
	Rollout Buffer Size (n_steps)	4096
	Entropy Coefficient (ent_coef)	3×10^{-5}
ICM Body Schema Model Parameters		
	VAE Latent Dim (Proprioception)	64
	VAE Latent Dim (Touch)	24
	VAE / Forward Model Hidden Dim	512
	Optimizer Learning Rate	3×10^{-4}
	VAE β Coefficient	0.01
Dynamic Weight Reward Mechanism Parameters		
	Stop Step for Weight Schedule	1 000 000
	ICM Reward Weight (λ_{icm})	$0.005 \rightarrow 0.1$ (Linear)
	Body Touch Reward Weight (λ_{touch})	$10.0 \rightarrow 2.5$ (Linear)
	Hand Touch Reward Weight (λ_{hand})	$80.0 \rightarrow 8.0$ (Linear)
	ICM Training Epochs (n_epochs)	2
	ICM Training Batch Size (batch_size)	256

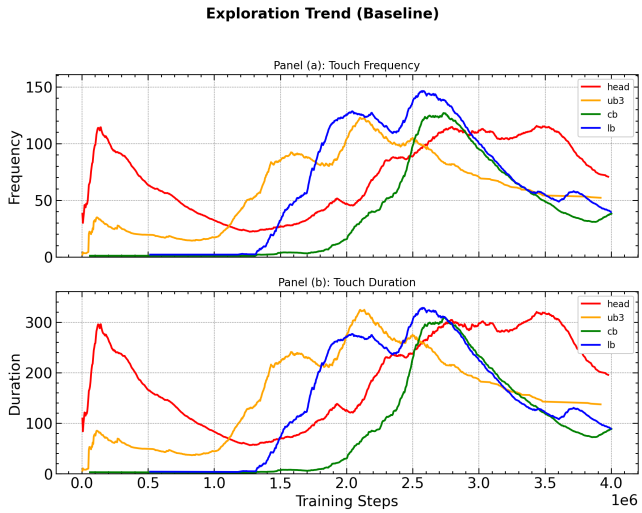


Fig. 3: Evolution of exploration focus over time for the baseline model. The plot shows the smoothed frequency (Panel a) and duration (Panel b) of touches for key body parts, revealing a sequential, rostro-caudal trend.

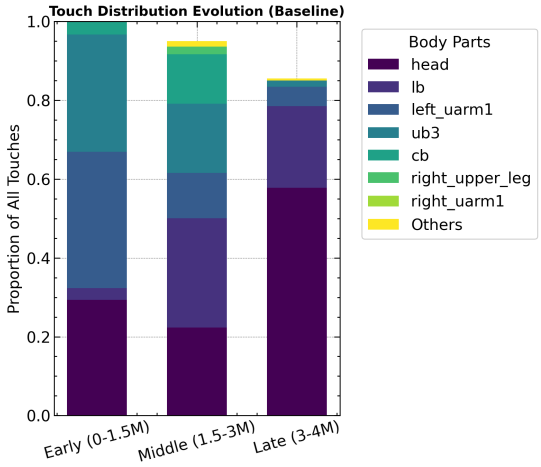


Fig. 4: Evolution of the touch distribution for the baseline model across three developmental stages, showing a shift to a more comprehensive, full-body exploration.

phase of "skill acquisition" (learning *how* to touch) to one of "knowledge acquisition" (learning the *consequences* of touch).

2) Ablation Study: Dissecting the Intrinsic Motivation Components: The superiority of the complete framework is evidenced by the exploration diversity comparison in Fig. 7, where the **Baseline** model significantly outperforms all ablation conditions. The ablation studies systematically reveal the critical function of each component.

The **No Touch Reward** experiment demonstrates that the touch reward serves as a necessary **behavioral anchor**. Driven by pure, unfocused curiosity, the agent's exploration remains aimless and sporadic. Consequently, the touch distribution fails to mature beyond a nascent upper-body focus, and the internal model's learning ultimately fails. This confirms that pure curiosity is insufficient to guide the agent toward meaningful body schema learning. The results are detailed in Fig. 8.

Conversely, the **No Curiosity** experiment reveals that curiosity is the core **driver for sustained exploration**. As shown in Fig. 9, the agent exhibits "behavioral fatigue": after an initial period of exploration driven by the novelty of self-touch, the total frequency and duration of touches decline sharply. Lacking an epistemic drive to improve its world model, the agent loses motivation once the basic skill of self-touch is acquired.

The **Simple Touch Reward** experiment provides a compelling demonstration that a sophisticated, **habituation-mimicking mechanism is crucial for avoiding behavioral traps**. With a naive, fixed touch reward, the agent quickly falls into a "reward hacking" loop. It exhibits pathologically repetitive, high-frequency touches that are overwhelmingly concentrated on the most easily accessible body parts, completely failing to explore the rest of its body. This result powerfully

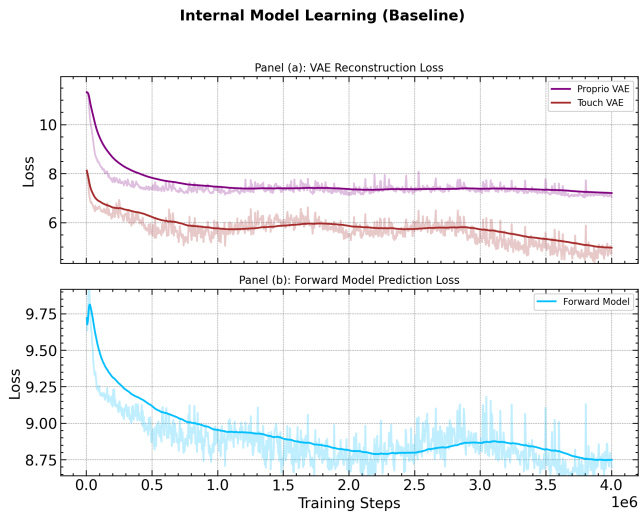


Fig. 5: Learning curves of the internal body schema model. Both VAE reconstruction losses (a) and the forward model prediction loss (b) exhibit a continuous downward trend.

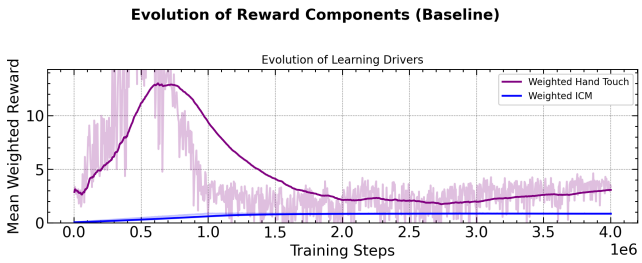


Fig. 6: Evolution of the weighted reward components, illustrating a developmental shift in learning drivers from touch-driven to curiosity-driven.

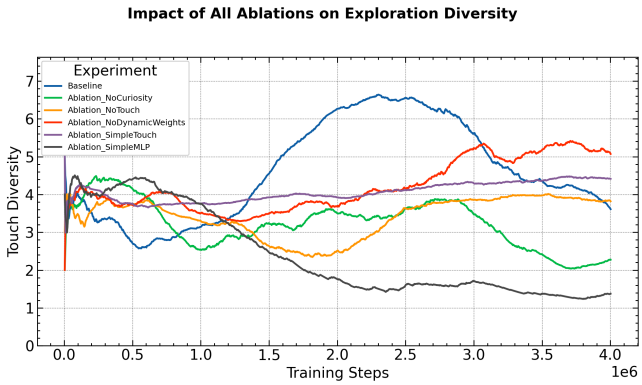
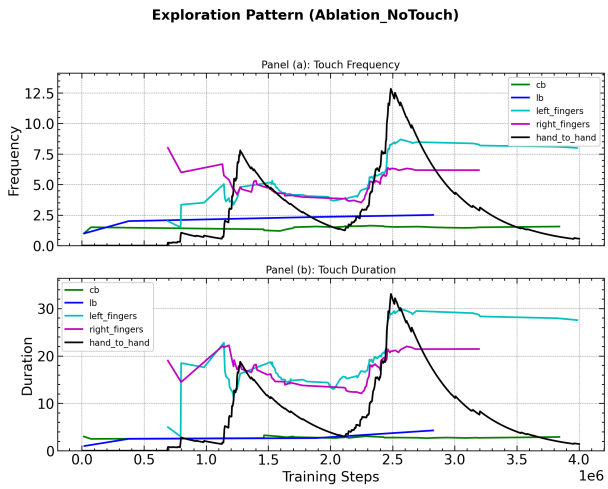
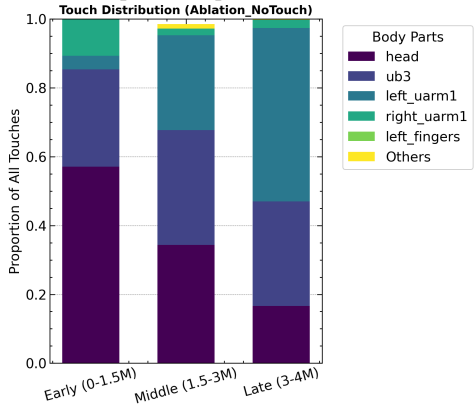


Fig. 7: Comparison of exploration diversity across all experiments. The baseline model consistently achieves a higher level of diversity than all five ablation studies.

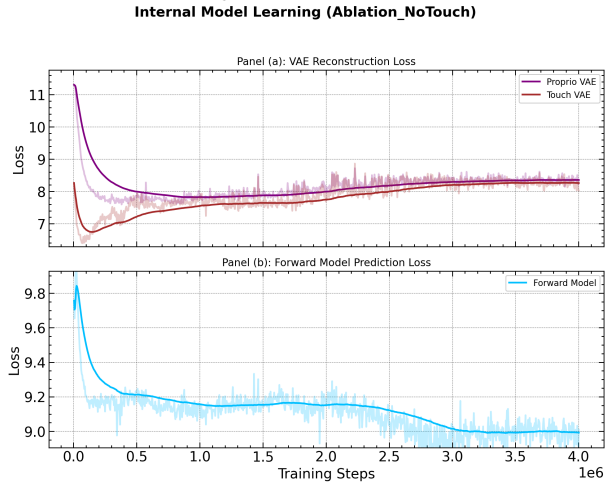
validates that our dynamic, habituation-based reward is vital for encouraging diverse and comprehensive exploration, rather than myopic exploitation of the easiest reward source (see Fig. 10).



(a) Aimless exploration pattern without touch reward.



(b) Stagnant touch distribution.



(c) Deteriorating VAE reconstruction loss.

Fig. 8: Results for the **No Touch Reward** ablation. The agent exhibits aimless exploration (a), a failure to develop a comprehensive touch distribution (b), and deteriorating internal model performance (c).

3) Ablation Study: Validating the Curriculum and Representation Model: Our results further show that both the devel-

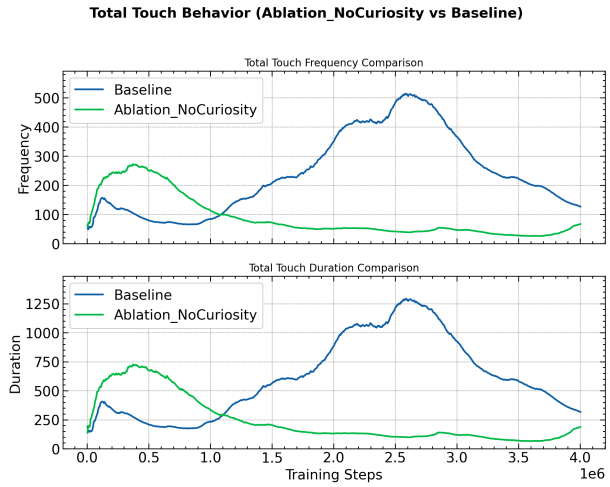
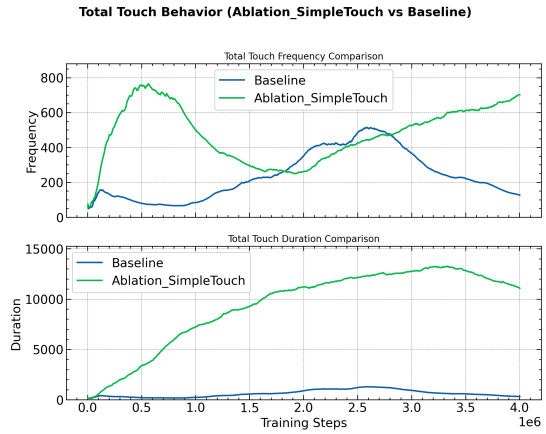
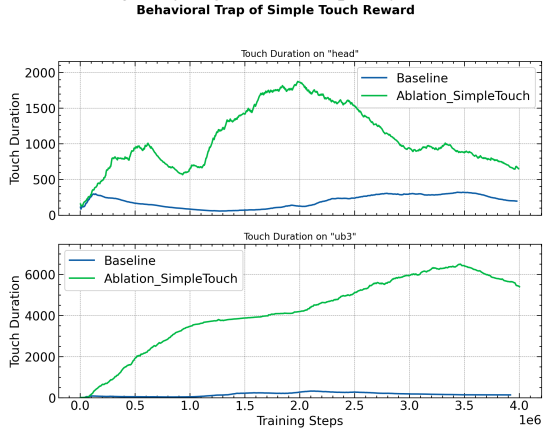


Fig. 9: Total touch behavior for the **No Curiosity** agent. The agent exhibits a "behavioral fatigue" pattern, with activity declining sharply after a peak.



(a) Pathologically high touch frequency and duration.

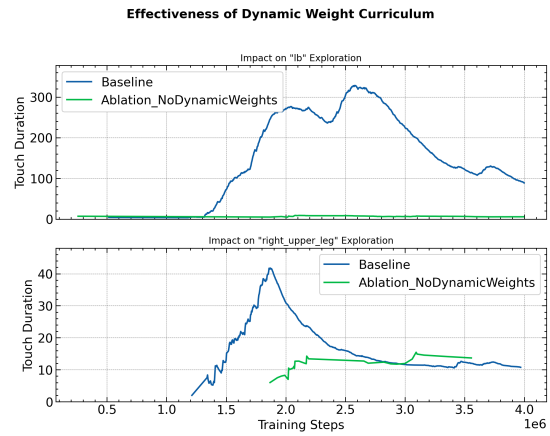


(b) Behavioral trap with fixation on head and upper body.

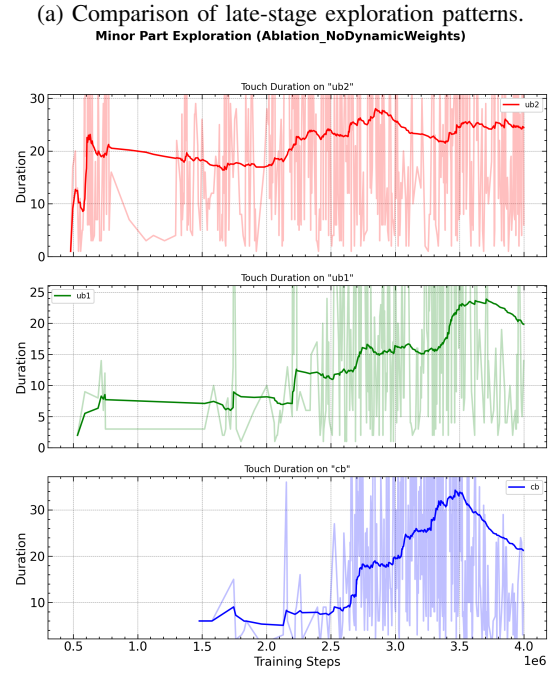
Fig. 10: Results for the **Simple Touch Reward** ablation. The agent falls into a "reward hacking" trap, indicated by abnormally high touch activity (a) and a behavioral fixation on easily accessible body parts (b).

opmental curriculum and the quality of the state representation are indispensable.

The **No Dynamic Weights** agent, which was trained without the curriculum, failed to master the more complex, late-stage exploratory skills. As seen in Fig. 11, while its overall performance was reasonable, it did not establish the robust and sustained exploration patterns for the lower back (`lb`) and right leg that were characteristic of the baseline model. This demonstrates that the curriculum, which prioritizes skill learning before knowledge building, is **critical for bootstrapping the acquisition of more advanced behaviors**.



(a) Comparison of late-stage exploration patterns.

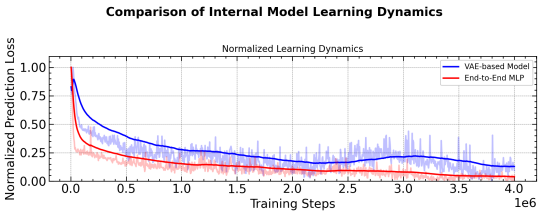


(b) Sparse and unstable exploration of minor body parts.

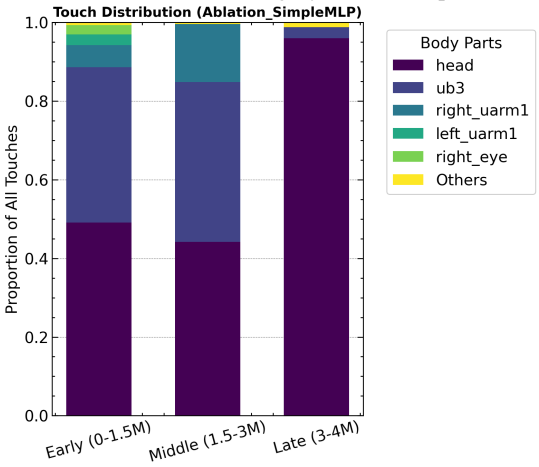
Fig. 11: Results for the **No Dynamic Weights** ablation. Without the curriculum, the agent fails to develop robust exploration of the lower body (a) and its exploration of other parts remains sparse (b).

Finally, a high-quality state representation is the **cornerstone of the entire framework**. The agent in the **Simple MLP**

Model ablation, which lacked the VAE representation, suffered a catastrophic failure. While its prediction loss appeared to decrease, its behavior collapsed into a single, repetitive action: touching its head, which accounted for over 95% of all contacts in the late stage. We diagnose this as a **"Curiosity Trap"**: the inefficient MLP struggles to effectively model even this simple action, causing the prediction error to remain persistently high. This, in turn, provides a constant, inescapable curiosity reward that locks the agent into a single behavior. This outcome powerfully demonstrates that an efficient and robust state representation model, such as our VAE, is a fundamental prerequisite for ensuring that curiosity-driven exploration remains effective and avoids such pathological behavioral traps (see Fig. 12).



(a) Internal model learning dynamics comparison.



(b) Behavioral collapse into a single action.

Fig. 12: Results for the **Simple MLP Model** ablation. While the MLP's loss decreases (a), its inefficient representation leads to a "Curiosity Trap," causing behavior to collapse into repetitively touching the head (b).

V. Discussion

This work presented a developmentally-inspired computational framework to investigate a core question in cognitive science: how can an agent, without external task guidance, autonomously learn and construct an effective body schema through intrinsically motivated self-exploration? Our experimental results not only demonstrate the viability of this framework but also offer profound insights into the mechanisms underlying early cognitive development in infants.

A. Theoretical Contributions

First and foremost, our model puts forth a **concrete, computable hypothesis** for the fundamental problem of how infants build body knowledge through unsupervised, intrinsically motivated processes. In contrast to traditional views of development as a series of pre-programmed maturational stages, our work demonstrates how complex, ordered behavioral patterns can **emerge** from simple, local learning rules within a self-organizing system. Specifically, our model, without any spatial priors, autonomously replicated the "rostro-caudal" (head-to-tail) progression of touch exploration that is highly consistent with findings in human infants [6], [27]. This result strongly supports the theory that the seemingly random exploratory actions of early infancy may, in fact, constitute a highly efficient process for body schema construction, guided by internal cognitive mechanisms such as curiosity and novelty preference, as modeled here. The body model established through such exploration likely serves as the cognitive foundation upon which more complex, transferable motor skills are subsequently developed.

Second, within the field of developmental robotics, our work demonstrates that an agent driven purely by intrinsic curiosity can successfully construct its own body model. Unlike competence-based or goal-directed paradigms such as Goal Babbling [7], [21], which often rely on a predefined goal space, our framework employs a more open-ended and fundamental knowledge-based exploration strategy. By rewarding "learning progress" in its world model (for which prediction error serves as a proxy), our agent exhibits more flexible and diverse exploratory behaviors. This not only validates the utility of developmental robotics as a tool for conducting "thought experiments" on human cognition [1] but also presents a viable pathway toward creating more autonomous and adaptive general-purpose robots.

B. Limitations and Future Work

Despite the promising results, this study has several limitations, which in turn open up new avenues for future research.

First is the **sensitivity to hyperparameters**. We observed that the model's performance is somewhat sensitive to the reward weights and the parameters of the dynamic curriculum. Although the current settings produce stable and developmentally plausible behaviors, they were found through manual tuning and may not be optimal. Future work could explore methods for automated curriculum learning [28], allowing the agent to adaptively schedule the weights of its motivations based on its own learning progress.

Second is the **absence of temporal memory and attention**. The current agent's policy and body schema model are reactive, based solely on the current state. Human learning, however, is deeply reliant on memory and attention. Integrating recurrent architectures (e.g., RNNs or Transformers) into the model, enabling it to process a history of sensorimotor experiences, could allow for the emergence of more complex behavioral patterns, such as the developmentally-evolving

”complex touches” (continuous sliding motions) observed by DiMercurio et al. [6].

Third, a promising direction is to enable **hierarchical learning from exploration to skills**. Our current model successfully demonstrates the emergence of exploratory behaviors but does not consolidate them into explicit, reusable motor skills. A compelling next step would be to abstract the learned self-touch ”motor primitives” as ”options” within a Hierarchical Reinforcement Learning framework [29]. Higher-level intrinsic motivations, such as competence improvement [19], could then be introduced to drive the agent to learn how to compose these options to achieve more complex tasks, such as stably grasping its own foot.

Fourth, there are **limitations in the simulation's fidelity**. While the MIMo platform provides a high-fidelity physical simulation, the granularity of the tactile data is still coarse. We can detect contact between body geometries, but cannot distinguish finer touch modalities (e.g., palmar vs. dorsal contact) or complex actions like ”grasping.” This limits our ability to make direct comparisons with more nuanced infant studies [5]. Future work involving higher-resolution tactile sensing or transferring the model to a physical robotic platform will be crucial for further validation and extension.

VI. Conclusion

In this paper, we proposed a unified, developmentally-inspired computational framework that enables a robotic agent to successfully learn its multimodal body schema through intrinsically motivated self-exploration, devoid of any external rewards or predefined goals. By integrating a probabilistic, generative body model, a prediction-error-based curiosity module, and a dynamic touch reward mechanism that simulates infant habituation and interest shift, our model autonomously developed ordered exploratory patterns from random movements that are remarkably consistent with those of human infants. Our research not only provides a concrete, computable model for the core mechanisms of early cognitive development but also contributes new methods and insights to the field of developmental robotics for building more autonomous and adaptive intelligent agents.

References

[1] A. Cangelosi and M. Schlesinger, “From babies to robots: the contribution of developmental robotics to developmental psychology,” *Child Development Perspectives*, vol. 12, no. 3, pp. 183–188, 2018.

[2] F. De Vignemont, “Body schema and body image—pros and cons,” *Neuropsychologia*, vol. 48, no. 3, pp. 669–680, 2010.

[3] M. Hoffmann, H. Marques, A. Arieta, H. Sumioka, M. Lungarella, and R. Pfeifer, “Body schema in robotics: a review,” *IEEE Transactions on Autonomous Mental Development*, vol. 2, no. 4, pp. 304–324, 2010.

[4] P. Rochat, “Self-perception and action in infancy,” *Experimental brain research*, vol. 123, no. 1, pp. 102–109, 1998.

[5] 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.

[6] 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.

[7] M. Rolf, J. J. Steil, and M. Gienger, “Goal babbling permits direct learning of inverse kinematics,” *IEEE Transactions on Autonomous Mental Development*, vol. 2, no. 3, pp. 216–229, 2010.

[8] 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.

[9] P.-Y. Oudeyer and F. Kaplan, “What is intrinsic motivation? a typology of computational approaches,” *Frontiers in neurorobotics*, vol. 1, p. 108, 2007.

[10] P. D. Nguyen, Y. K. Georgie, E. Kayhan, M. Eppe, V. V. Hafner, and S. Wermter, “Sensorimotor representation learning for an “active self” in robots: a model survey,” *KI-Künstliche Intelligenz*, vol. 35, no. 1, pp. 9–35, 2021.

[11] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.

[12] S. Rezaei-Shoshtari, F. R. Hogan, M. Jenkin, D. Meger, and G. Dudek, “Learning intuitive physics with multimodal generative models,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 7, 2021, pp. 6110–6118.

[13] R. Grush, “The emulation theory of representation: Motor control, imagery, and perception,” *Behavioral and brain sciences*, vol. 27, no. 3, pp. 377–396, 2004.

[14] D. Pathak, P. Agrawal, A. A. Efros, and T. Darrell, “Curiosity-driven exploration by self-supervised prediction,” in *International conference on machine learning*. PMLR, 2017, pp. 2778–2787.

[15] J. Schmidhuber, “Formal theory of creativity, fun, and intrinsic motivation (1990–2010),” *IEEE transactions on autonomous mental development*, vol. 2, no. 3, pp. 230–247, 2010.

[16] R. L. Fantz, “Visual experience in infants: Decreased attention to familiar patterns relative to novel ones,” *Science*, vol. 146, no. 3644, pp. 668–670, 1964.

[17] V. Dumont, J. Bulla, N. Bessot, J. Gonidec, M. Zabalia, B. Guillois, and N. Roche-Labarbe, “The manual orienting response habituation to repeated tactile stimuli in preterm neonates: discrimination of stimulus locations and interstimulus intervals,” *Developmental Psychobiology*, vol. 59, no. 5, pp. 590–602, 2017.

[18] J. Bongard, V. Zykov, and H. Lipson, “Resilient machines through continuous self-modeling,” *Science*, vol. 314, no. 5802, pp. 1118–1121, 2006.

[19] 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.

[20] K. Zhang, P. Yun, J. Cen, J. Cai, D. Zhu, H. Yuan, C. Zhao, T. Feng, M. Y. Wang, Q. Chen *et al.*, “Generative artificial intelligence in robotic manipulation: A survey,” *arXiv preprint arXiv:2503.03464*, 2025.

[21] F. Gama, M. Shcherban, M. Rolf, and M. Hoffmann, “Goal-directed tactile exploration for body model learning through self-touch on a humanoid robot,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 15, no. 2, pp. 419–433, 2021.

[22] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, “Stable-baselines3: Reliable reinforcement learning implementations,” *Journal of machine learning research*, vol. 22, no. 268, pp. 1–8, 2021.

[23] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, “Curriculum learning,” in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 41–48.

[24] F. M. López, V. Marcel, X. Hinaut, J. Triesch, and M. Hoffmann, “Babybench: A multimodal benchmark of infant behaviors for developmental ai,” 2025.

[25] E. Todorov, T. Erez, and Y. Tassa, “Mujoco: A physics engine for model-based control,” in *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2012, pp. 5026–5033.

[26] D. Mattern, P. Schumacher, F. M. López, M. C. Raabe, M. R. Ernst, A. Aubret, and J. Triesch, “Mimo: A multimodal infant model for studying cognitive development,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 16, no. 4, pp. 1291–1301, 2024.

[27] J. Khoury, S. T. Popescu, F. Gama, V. Marcel, and M. Hoffmann, “Self-touch and other spontaneous behavior patterns in early infancy,” in *2022 IEEE International Conference on Development and Learning (ICDL)*. IEEE, 2022, pp. 148–155.

[28] S. Narvekar, B. Peng, M. Leonetti, J. Sinapov, M. E. Taylor, and P. Stone, “Curriculum learning for reinforcement learning domains: A framework and survey,” *Journal of Machine Learning Research*, vol. 21, no. 181, pp. 1–50, 2020.

[29] A. G. Barto, S. Singh, N. Chentanez *et al.*, “Intrinsically motivated learning of hierarchical collections of skills,” in *Proceedings of the*

3rd International Conference on Development and Learning, vol. 112.
Piscataway, NJ, 2004, p. 19.