# Tracking Action in XR Mark Test Reveals Developing Body Representation Among Toddlers

Michiko Miyazaki[1*], Tomohisa Asai[2], Norihiro Ban,[3] and Ryoko Mugitani[4]

[1] Faculty of Social Information Studies, Otsuma Women's University, Tokyo, 102-8357
[2] Department of Cognitive Neuroscience, ATR (Advanced Telecommunications Research Institute International), Kyoto, 619-0288, Japan
[3] HEIMEI Corporation, Osaka, 564-0051, Japan
[4] The Faculty of Integrated Arts and Social Sciences, Japan Women's University, Tokyo, 112-0015, Japan

*Michiko Miyazaki
Email: myzk@otsuma.ac.jp

Acknowledgments
We thank Ayano Ryoke and Moe Kobayashi for their help with the data collection and analysis. We also thank all caregivers and children who participated in this study. This work was supported by JSPS KAKENHI (Grant Numbers 19H04019, 23H03704) to Michiko Miyazaki, Tomohisa Asai, and Ryoko Mugitani.

Author Contributions:
All authors contributed to the study concept, design, and interpretation. N.B programmed the Bodytoypo task, M. M. collected the data, T.A. developed and performed the analyses, M. M., T. A., and R. M. wrote the manuscript. All authors approved the final version of the manuscript for submission.
Competing Interest Statement:
The authors declare no competing interests.

**Abstract**

The mark test, a popular index for self-recognition, has not been widely used as a paradigm for assessing self-body representations. This study demonstrates novel aspects of children's body representation and its development by combining the classic mark test with a cross-reality task, where participants' skeletal data were detected in real-time and virtual marks were displayed on 30 body parts, which participants could "touch." The accuracy and trajectory of the first touch and reaction time were analyzed. Thirty 2- and 3-year-olds participated in the task, and their parents reported on their bodily word acquisition. Touch errors at joints and movable body parts decreased from 2 to 3 years, regardless of vocabulary richness. Localization error could be predicted by dynamical body part coordination. Three-year-olds displayed fast and predictive (feedforward) reaching instead of visually guided (feedback) careful reaching. Analyzing hand-reaching strategies in the XR mark test revealed the development of sensorimotor body representations.

*Keywords*: mark test; body representation; OpenPose; feedback control; feedforward control

## Statement of Relevance

In this study, the mark test was extended to a body part localization XR task called Bodytoypo. By combining real-time skeletal detection and augmented reality, participants can engage in the task intuitively, without language use. This suggests that Bodytoypo could be applied to young children, adults, and other animal species with appropriate updates. Furthermore, by incorporating automated machine learning methods in analyzing body part localization errors and reaching strategies, developmental changes in reaching patterns can be objectively demonstrated, and changes in weighting across multiple reaching strategies can be evaluated. This reduces the arbitrariness of manual classification by human observation. The action-based evaluation also opens new avenues for approaching the general question of self-recognition. This project is not simply a digitization of the mark test but a rich collaboration of engineering, machine learning statistics, and developmental psychology.

## Introduction

The mark test, developed by G. Gallup in 1970, evaluates *self-recognition*, and has been studied extensively for the past 50 years among various species (Amsterdam, 1972; Anderson & Gallup, 2015; Gallup, 1970; Lewis & Brooks-Gunn, 1979; Suddendorf & Butler, 2013). In the test, participants discover a mark on their face through their mirror image and spontaneously touch or attempt to remove it. However, passing the test as an indicator of self-recognition is still controversial (Anderson & Gallup, 2015; de Waal, 2019; Rochat & Zahavi, 2011; Suddendorf & Butler, 2013).

We suggest reinterpreting the mark test as a “paradigm” that evokes self-oriented, visually-guided, and spontaneous reaching movements, inducing the emergence of a sense of body ownership and agency. The image of one's body in a mirror or video becomes the objectified self for toddlers who have just acquired mirror self-recognition. The acquisition of the recognition and manipulation of the body reflected in mirrors and videos, for example, helps evaluate the emergence of a sense of body ownership and agency, as examined in the rubber hand/robot hand illusion (RHI).

Toddlers’ reaching errors effectively reflect developmental changes in body representations. For example, in our previous study, approximately 40% of 2-year-olds found a mark placed on their forehead but initially tried to remove it from the back of their heads (the rear-search error (Imaizumi et al., 2021; Miyazaki & Hiraki, 2017). This error may result from the immaturity and entanglement of cognitive abilities related to the acquisition of body representations, including (1) the acquisition of the meaning and relative associations between body parts (body representation), (2) multimodal integration with visual information, and (3) motor control of the arm used to reach (acquisition of internal models). By analyzing the cognitive background of reaching errors from these three properties, we can highlight the development of body representations from multiple perspectives as follows.

First, the aspect of the mark test to evoke self-oriented behavior involves acquiring meanings and relative associations between body parts (semantic body representation (Schwoebel & Coslett, 2005)). For example, the upper forehead has no independent body name and may have a less clear relative positional relationship than the nose, where rear-search errors never happen. Several studies show that the speed of name acquisition and the difficulty of body part pointing vary by body part (Brownell et al., 2007, 2010; Camões-Costa et al., 2011; Waugh & Brownell, 2015; Witt et al., 1990). To examine the relationship between body part name acquisition and reaching behavior, we analyzed the relationship between questionnaire answers on vocabulary acquisition status and reaching for marks. We aimed to clarify the differences in localization difficulty and coordination by body part to elucidate the so-called *body map* or *sensory homunculus*.

Second, the specificity of children's visual processing can be examined in terms of how they integrate visual (mirrored body) and interoceptive information (e.g., proprioception, touch). Rear-search errors may result from a misalignment of the real body and the mark owing to the priority given to the visible mirrored body. Moreover, studies of RHI in children indicate that it may require significant time for the integrated processing of vision, touch, and proprioceptive senses (Gottwald et al., 2021).

However, RHI among younger children is considered difficult to comprehensively examine because it requires verbal skills for subjective reports and fine motor control that indicates proprioceptive drift (Cowie et al., 2013, 2016). In contrast, reaching a visual target is a multisensory processing task applicable to children. Shen et al. (2010) found that 1- to 5-year-olds visually perceive a target by stabilizing the head before reaching, resulting in smooth reaching. Visually guided reaching integrates the body and mirror image as a manipulation target from the inside and outside. Clarifying how children master reaching is important for understanding the learning mechanism of multisensory processing related to acquiring body

representations.

Finally, the property of spontaneous reaching, tied to visually guided reaching, can reveal developmental changes and changes in weightings in the internal model. The rear-search error may be caused by an attempt to map the mirror image (vision) to the real body (proprioception) through feedforward (FF) control to further localize the mark. Previous studies suggest that acquisition of reaching proceeds from feedback (FB) control, gradual visual-motor integration, and FF control acquisition with prediction (Bernstein, 1967; Scott Kelso, 2014). However, even newborns could reach for nearby targets if their trunks were supported and their posture was stable (Hofsten & von Hofsten, 1984), despite not seeing their hands (Claxton et al., 2003; Corbetta et al., 2014, 2018). Examining the developmental process of FB and FF control weighting in reaching, in which each body part is a target, we can clarify the development of the internal model and the dynamics of behavioral control.

Therefore, the three properties of the mark test can be considered a paradigm that illuminates new aspects of the mark test as a body part localization task. The mark test provides insight into how visual information and prediction are incorporated to achieve correct localization and how each body part is coordinated to touch the target on body parts. The three properties listed here are intricate and difficult to examine individually. However, all are essential for capturing developmental changes in body representations. In 2019, we applied the mark test to develop a body part localization task that spontaneously and repeatedly elicits reaching an invisible body part, the face (Miyazaki et al., 2019). In this study, we extend the previously developed task to a whole-body version (see Methods, Fig. 1 and Fig. S1A).

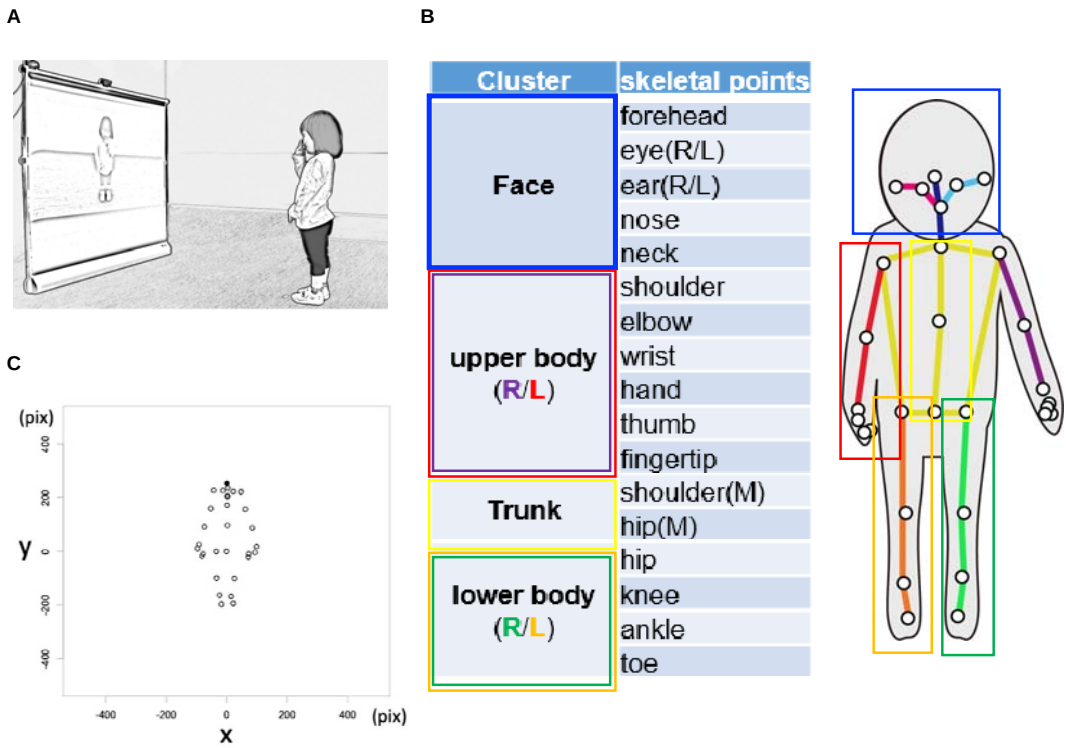


**Fig. 1.**

*Real-time Estimation of Body Parts for the XR Mark Test ("Bodytoypo").*

*Note*. (A) Online presentation for participants touching the AR mark in this study. (B) The original input image (frame of the video) with the estimated bone (30 parts) and six body clusters using automated AI analysis. R/L/M in the table indicate right, left, and midline, respectively. (C) The depicted output for x-y coordinates of each body part (30 fps) for the following offline motion analysis. The location of the center of the hip (Chip, 22) was always fixed at the origin (0,0) as the reference point.

By integrating skeletal detection with Openpose (Cao et al., 2021) and augmented reality (AR) technology, virtual marks were displayed on 30 parts of the participants' bodies. Their localization errors were comprehensively analyzed. To analyze reaching movement, the reaching motion was divided into latency and duration according to the change point. Bodytoypo is an interactive *XR* mark test that crosses the boundary between real and virtual, where participants can "touch" the AR marks.

## Methods

### Participants

Thirty-six Japanese children aged 2.5–3.5 years old participated. The sample size was determined based on previous studies on body representations (Camões-Costa, et al., 2011; Somogyi, et al., 2018). This study was conducted under the recommendations of the Otsuma Women's University Life Sciences Research Ethics Committee, and written informed consent was obtained from all the participants' parents. The study protocol was also approved by the Otsuma Women's University Life Sciences Research Ethics Committee (2019-012). The final sample comprised 30 children (mean = 34.7 months, range: 28–44 months, SD = 4.49), 18 2.5-year-olds (11 females), and 12 3.5-year-olds (9 females). Six participants were excluded from the analysis (attrition rate = 16.7%) owing to fussiness and shyness (n = 5), or issues with the machine (n = 1).

The final dataset used for AI-statistics analysis comprised 607 (2.5-years-old, n = 310; 3.5-years-old, n = 297) of the 852 manual trials of 30 children. Of these, 245 trials were excluded from the AI analysis (28.7% attrition) owing to several reasons. The main reason for omission was that the reaction time was too long (over 3s is approximately 3SD + mean), while other omissions were due to the children not noticing the mark, having little motivation for touch, or wandering around and playing. Trials that were deemed inappropriate for motion trajectory analysis were omitted.

## Apparatus

We developed a body part localization task using AR with the image processing library, Openpose. We named this system *Bodytoypo*. An image of each participant's whole body was recorded using a USB camera (Logicool C920) and displayed via a projector (Epson EB 485WT) onto a screen (KIMOTO RUM60N1) (see Fig. 1: it was presented in a mirror-like or ipsilateral relationship). The 2D coordinates of their body parts were automatically detected using Openpose on a GPU machine (Mouse computer NEXTGEAR i690PA2). The detected skeletal data were processed in real-time so that the AR character (favored cartoon characters of children) was displayed as a mark on the coordinates of the target body parts. The net delay (transmitting and image processing) was approximately ten frames (0.33 s), and the processing speed of the AR overlay was 15–22 fps (best-effort format). Participants stood on a sign 115 cm away from the screen. The camera angle was adjusted so that the entire body was reflected on the screen. Each participant's behavior was recorded using a capture device (Avermedia AVT-C878) on a stand-alone PC (Galleria QSF1070HGS). The recorded video was further used for offline analysis manual coding and movement analysis.

## Tasks

### *Body Part Localization Task (Bodytoypo)*

The experimenter manually displayed the AR mark on the participant's life-size self-image, and observed whether the participant could correctly touch (or point to) the actual body part corresponding to the mark's location. The AR marks were displayed individually on 30 body parts (upper forehead, forehead, eye (right/left, hereafter R/L), ear (R/L), nose, cheek (R/L), mouth, chin, collarbone (R/L), shoulder (R/L), upper arm (R/L), elbow (R/L), navel, forearm (R/L), hand (R/L), thigh (R/L), and big toe (R/L)). Note that the 30 body parts

shown in Bodytoypo were based on the Openpose definition, which slightly differed from those in the independent offline analysis (see Fig. 1B, Fig. S1B, C). The order of the marks was randomly arranged in a sequence to avoid adjacent body parts. These sequences were presented in forward and reverse directions. The presentation order was counterbalanced among participants. If the participants could touch the correct body part with the AR mark, the mark disappeared with a fun audio-visual reward presented by the experimenter's manual keyboard control.

### *Questionnaire of Body Part Vocabulary*

After completing the task, parents were asked about their children's vocabulary acquisition status through a questionnaire. They were asked about their speech and comprehension of 60 body parts and body parts in relation to people and animals (e.g., tail, wings, horns) in Japanese.

## Procedure

The experimenter first explained the study to the parents to obtain informed consent. The experimenter and participating children built rapport through free playing. The experimenter further demonstrated ten trials of Bodytoypo to the participants. During the demonstration, she intentionally touched the mark incorrectly and indicated that if they made a mistake, the reward would not appear. When participants were suitably positioned by standing straight and still in front of the screen, an AR mark would appear with a beep sound. The experimenter encouraged them to participate in the task by rhythmically presenting the AR marks. If the participants did not readily respond to the presented mark, the experimenter re-presented it or skipped the trial. If participants lost motivation, trials of touching the nose, which was easy to touch, was inserted as needed to recapture motivation. Aside from those additional nose trials, a maximum of 30 trials (30 body parts) was executed for each participant. The experiment lasted five minutes.

**Quantification and Statistical Analysis**

The study analysis focused on the period between the appearance of the mark and the children's first touch in each trial. A detailed analysis of the accuracy and error of the first touch and the motor trajectory for each body part revealed developmental changes in the acquisition of body representations from the three perspectives presented earlier.

We employed both traditional frame-by-frame manual coding analysis and statistics-based automated analysis. The former analysis examined the number of trials performed, overall and body part-specific error rates (Fig. 2), and the relationship between Bodytoypo performance and body part-related lexical acquisition (Fig. 3). This manual analysis allowed us to examine whether the Bodytoypo task could sustain participants' motivation so that we could then show (1) the development of body representation. The latter analysis, statistics-based analysis, focused on (2) developmental changes in multimodal processing with visual information, and (3) developmental changes in internal models and their weighting through a detailed analysis of movement trajectories using 2D image processing.

The overall pipeline in the latter analysis is shown below. The necessary process was first; the change of point detection as each trial (from the mark appearing to the first touch) comprised two periods: response latency and reaching movement (Fig. 4). Therefore, the change point was estimated using the machine-learning R library “changepoint” (https://github.com/rkillick/changepoint/) over the time series of variance among distances between the used hand and other body parts (Fig. S4). During a reach, these distances change drastically compared with those during latency. This automated estimation (separation of latency and reaching) was further validated during a process conducted later (Fig. 5ABC and 6AE). Once we distinguished them, the whole-body associations were calculated as a temporal-correlation matrix during those two periods (Fig. 5), indicating sensorimotor coordinates for touching the AR mark. This summarized matrix exhibits a relative similarity of trajectory among body parts so that the original configuration (physical body) was partly

recovered from the matrix through multidimensional scaling (MDS) projection during latency (Fig. 6A) rather than during reaching (Fig. 6E).

Furthermore, this correlation matrix was vectorized and entered into the regression analysis (as inputs) to predict children's touch errors (as outputs) (Fig. 6BD). During each period, body part associations were visualized at a glance in terms of the x-y joint dendrogram (tanglegram, Fig. 6CF). Lastly, all the variables were put into a linear mixed-model (all-in-one statistical model) as continuous variables, to predict touch error (Fig. 7A). In summary, we established that 2- to 3-year-olds' sensorimotor representation can be depicted where the FB or FF weight for touching seems to be adaptively changing. Furthermore, the effectiveness of the automated analysis is evaluated by applying the findings from the statistics-based automated analysis to the manual analysis, from a qualitative perspective.

## Results

### Maintaining Participants' Motivation and Overall Error Rate

The average numbers of executed trials were 27.50 (SD = 5.371) and 29.75 (SD = .452) for the 2- and 3-year-olds, respectively. Most participants completed 30 trials, regardless of age ($t(28) = -1.439$, $p = .161$, *n.s.*, effect size $d = -.522$). This result is consistent with our previous study (26) and suggests that we could maintain participants' motivation during this whole-body version. Furthermore, the positive effect of gamification (i.e., producing an enjoyable animation and sound) was demonstrated (see Methods).

The overall error rate for the first touch was 39.9% (SD = .132) and 35.0% (SD = .095) for the 2- and 3-year-olds, respectively. The age groups differed insignificantly ($t(28) = 1.098$, $p = .282$, *n.s.*, effect size $d = .398$). Figure 2 summarizes the mean error rates for each age and body part (Fig. 2A) and in mark locations (midline, left side, and right side; see Fig. 2B). We examined the differences in mean error rates by integrating body parts by mark locations and found that the error rate for midline body parts was significantly lower

regardless of age. The main effect of age and the interaction between age and mark location were insignificant ($F(1,28) = .830$, $p = .370$, $MSe = .044$, *n.s.*, partial $\eta^2 = .029$, $F(2,56) = .954$, $p = .391$, $MSe = .019$, *n.s.*, partial $\eta^2 = .033$, respectively). However, the main effect of mark location was significant ($F(1,28) = 35.647$, $p < .01$, $MSe = .019$, partial $\eta^2 = .560$). Holm's multiple comparisons showed that the error rates in the midline location were significantly lower than those in other body parts, and that the left-right difference was insignificant (midline-left: $p < .01$, effect size $d = -1.635$; midline-right: $p < .01$, effect size $d = -1.338$; left-right: *n.s.*, effect size $d = .385$).

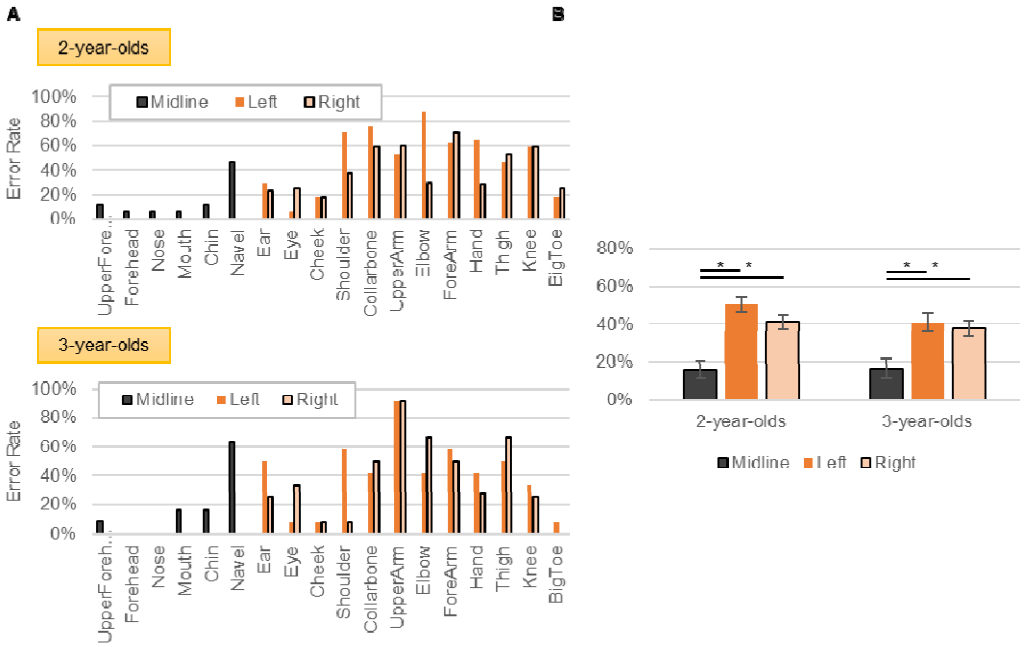


**Fig. 2.**

*Error Rates for all 30 Body Parts in Bodytoypo.*

*Note*. (A) Error rates for all 30 body parts by age. The body part around the face had a low error rate regardless of age. (B) Error rates for each age by position category (midline, left, and right). Error rates for body parts located on the midline were significantly lower than those for other body parts.

**Localization Errors and word acquisition**

We further analyzed the error rates of each body part combined with bodily function. The 30 body parts were re-categorized by bodily function and are summarized in Figure 3. A two-way ANOVA (age and bodily function) was used to explore whether the error rates differed. We examined four bodily functions, that is, whether the body part was a joint, moveable part, face, or named body part. The main effects of all bodily functions were confirmed (main effects of joint: $F(1,28) = 12.259$, $p = .002$, $MSe = .021$, partial $\eta^2 = .305$; moveable: $F(1,28) = 25.507$, $p < .01$, $MSe = .013$, partial $\eta^2 = .149$; face: $F(1,28) = 113.637$, $p < .01$, $MSe = .015$, partial $\eta^2 = .802$, and named body part: $F(1,28) = 58.481$, $p < .01$, $MSe = .017$, partial $\eta^2 = .676$. Additionally, the interaction of age and bodily function was significant regardless of whether it was a joint or a moveable part (age * joint: $F(1,28) = 4.898$, $p = .035$, $MSe = .021$, partial $\eta^2 = .149$; age * moveable: $F(1,28) = 4.026$, $p = .055$, $MSe = .013$, partial $\eta^2 = .126$). These results suggest that the accuracy of the first touch increases between the ages of 2 and 3 years for moveable joints and parts.

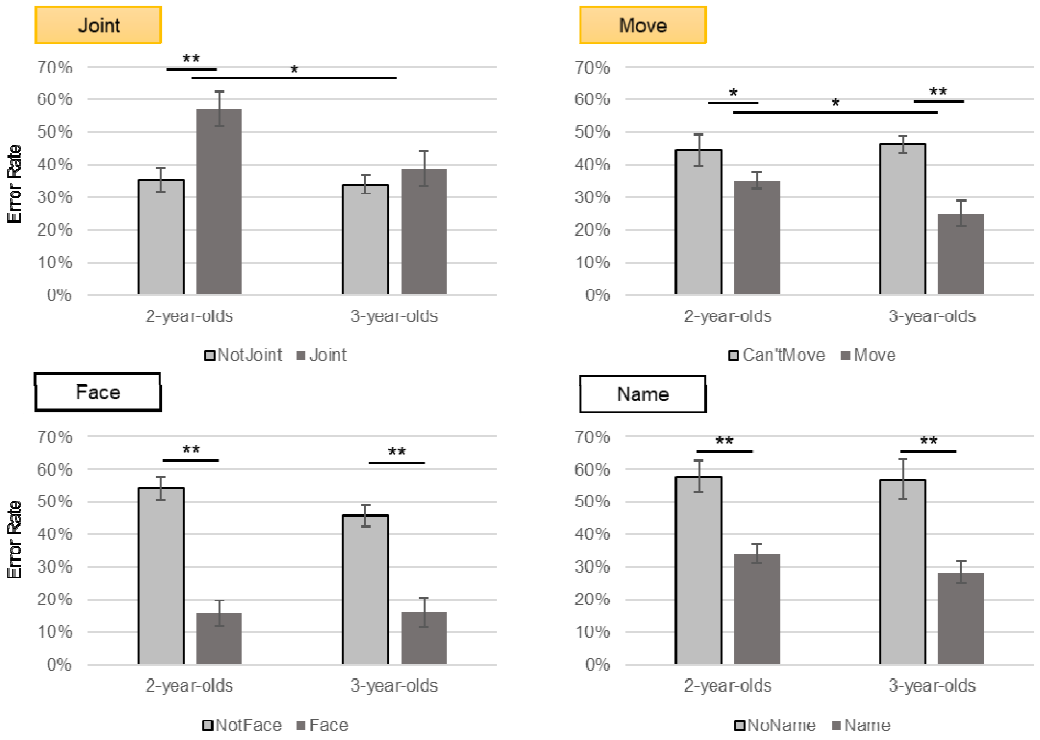


**Fig. 3.**

*Error Rates by Bodily Function*

*Note*. The 30 body parts presented by Bodytoypo were categorized into two groups from the following four perspectives, and the differences in error rates were examined. We used the following categories: joint or not, movable or not, face or not, and named or not. In all four categories, the main effects of categories were observed, but only for the categories of joint and movable, an interaction with age was observed.

The results of the body word acquisition questionnaire for parents revealed that for the 60 words asked, 28.7 (SD = 10.599) (2-year-olds) and 40.8 (SD = 14.640) (3-year-olds) of the words overall could be produced and comprehended (see Fig. S2C), respectively, There was a significant difference in the mean number of acquired words between the age groups ($t(28) = -2.632$, $p = .014$, effect size $d = -.955$). A partial correlation analysis was conducted on the number of acquired body vocabulary and Bodytoypo's error rate, with age in months as a control variable (Fig. S2A). A correlation analysis was also conducted between the acquisition rate for body parts used in Bodytoypo and the error rate (Fig. S2B). However, the results were insignificant.

## Visual-Guided Reaching Development

### *Body-Parts Association During Latency and Reaching*

Figure 4 summarizes the procedure of each trial and Figure 5 suggests the averaged temporal correlation among 30 body parts (29 parts except for the reference). The upper triangles represent x-, while the lower triangles represent y-coordinates where some "clusters" are observed on the diagonal (e.g., "face" cluster). The body parts (Fig. 5A) or clusters (parts-averaged, Fig. 5BC) were generally positively correlated, except for L-R upper body associations during the reaching period (blue squares). This negative correlation in the x-coordinate suggests bi-manual coordination in reaching behaviors, which could validate the change point estimation (see Fig. S4). Both age groups behaved similarly in terms of these correlation matrices (Fig. 5BC) as well as of the reaching trajectory in 2D (Fig. S5). However, the "face" (or head) behaved oppositely between 2- and 3-year-olds in the case of the subtracted matrix between correct and incorrect trials (yellow parts in Fig. 5D for reverse monochrome patterns) or the cross-correlation between the face and the hand (Fig. S7A). We observed what happened to those participants as follows.

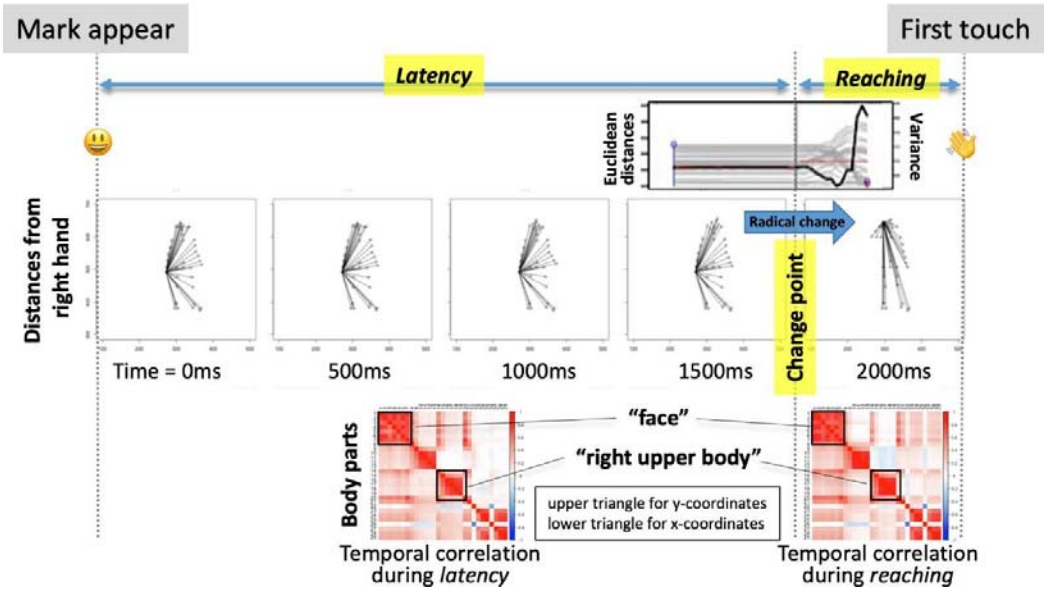


**Fig. 4.**

*Schematic Illustration for Each Trial*

*Note.* Each trial (from the mark appearing to the first touch) had two periods: response latency and reaching duration. This was defined by the detected change point over the time series of variance among distances between the used hand (in this example, right hand) and other body parts (the upper panel or supplementary Fig. S4).

During reaching, these distances dramatically changed compared with those during latency (middle successive panels). The association among body parts (temporal correlation of x- or y-components) was calculated for each period (lower two panels), where some "clusters" were observed on the diagonal (e.g., "face" cluster). This correlation matrix was further vectorized and used in the statistical analysis as inputs for predicting participants' touch errors as outputs.

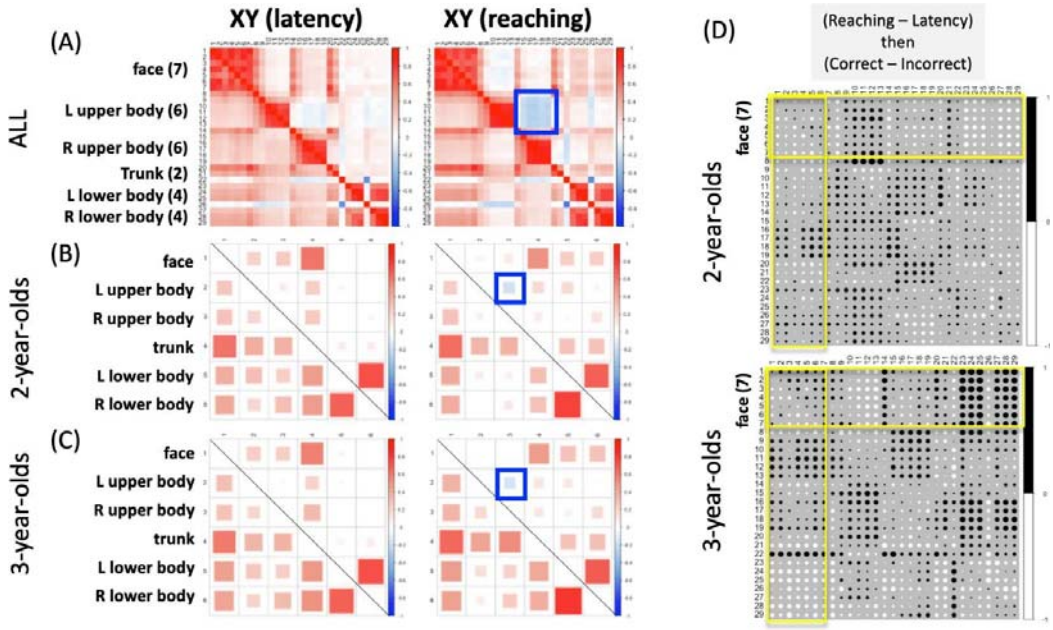


**Fig. 5.**

*Temporal Correlation Among Body Parts for 2- or 3-year-olds*

*Note*. Temporal correlation matrices for body parts during latency or reaching. The upper triangles are for x-, whereas the lower triangles are for the y-coordinates. Body parts (A) or clusters (B, C) were positively correlated in general, except for L-R upper-body relations during the reaching period (blue squares). This negative correlation in the x-coordinate suggests bimanual coordination in reaching behaviors that could validate the change point estimation. Both age groups behaved similarly in terms of the association among body clusters (B, C) as well as of the reaching trajectory (see Fig. S5). (D) However, the face behaved oppositely between 2- and 3-year-olds if we look at the subtracted matrix between correct and incorrect trials (see the main text for the detailed discussion).

***Correlation Matrix for Configuration, Prediction, and Coordination***

Figure 6A shows the recovered body configuration from the physical positions (averaged x-y coordinates of all trials, the upper panel) to the MDS projection (based on the averaged temporal correlation matrix among body parts and the lower-left panel) during latency. The comparison between them (through the Procrustes transformation in the lower right panel) suggests that the association between body parts during latency is relatively stable so that we can see the original body configuration. However, it is horizontally biased as the hand easily moved horizontally even during latency. This is not the case for reaching even where the average position was the same (Fig. 6E, upper panel). Although the correlation matrix was also markedly similar to that during latency (see Fig. 5A), the original body configuration was not recovered (Fig. 6E in the lower-left and right panels) because various reaching movements were included, which could re-validate the change point estimation.

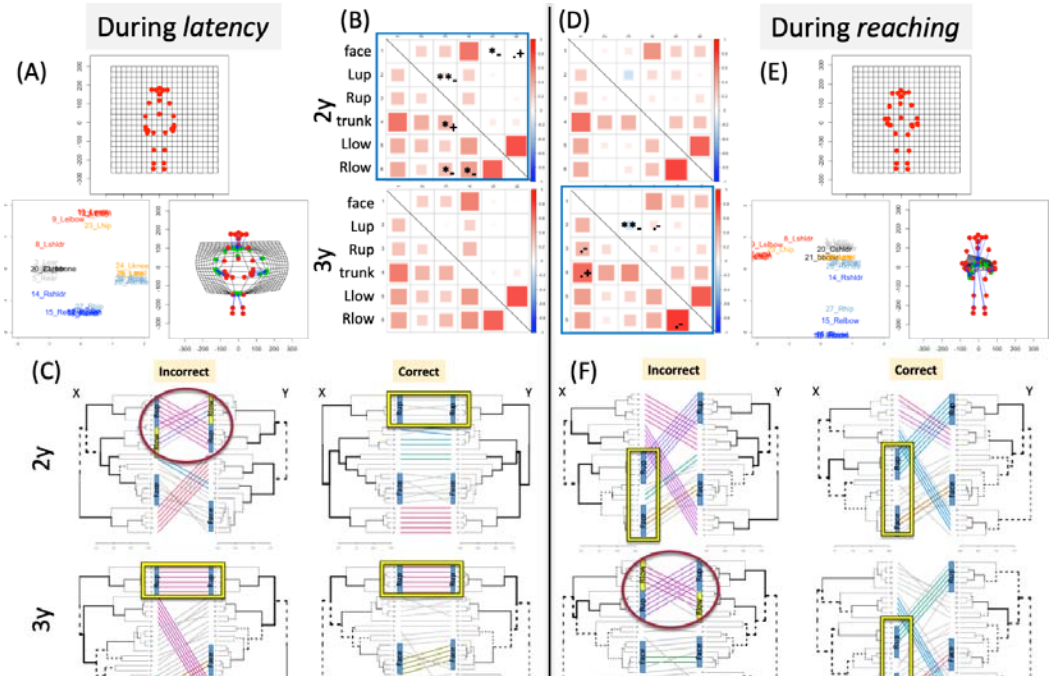

**Fig. 6.**

*Body Parts Associations for Configuration (AE), Prediction (BD), and Coordination (CF)*

*Note*. The temporal correlation matrices for latency (A) and reaching (E) were analyzed to recover the physical body-parts configuration through MDS projection and the Procrustes transformation. However, the reaching matrix did not globally recover the configuration due to the inclusion of various movements. The latency matrix for 2-year-olds can statistically predict the later touch errors (B) while the reaching matrix for 3-year-olds can predict the later touch errors (D). During the latency period, the incorrect tanglegram for 2-year-olds was characterized by a right upper and lower parts association (C). During reaching period, the same association was observed for the incorrect tanglegram for 3-year-olds (F). See main text for more discussion.

Although the correlation matrix between 2- and 3-year-olds was congruent (Fig. 6BD, reprinted from 5BC), there was a statistical difference in predicting their touch errors (overlaid asterisks). The correlation matrix was vectorized and entered into a linear regression model to predict the touch error distance. For 2-year-olds, the latency matrix as the assembled dataset (regardless of the participants or target positions at this stage) can predict touch error ($F(30, 436) = 1.666$, $p = .016$). By contrast, for 3-year-olds, the reaching matrix can predict touch error ($F(30, 308) = 1.636$, $p = .022$). However, the latency matrix for 3-year-olds or the reaching matrix for 2-year-olds cannot predict their touch errors ($F(30, 288) = 1.475$, $p = .057$; $F(30, 435) = 1.19$, $p = .229$, respectively). For each model, the significant (and marginal for reference) variables are indicated by symbols (Fig. 6BD, $p < .01$**, $p < .05$*, $p < .10$.) with signs (positive or negative effects), which include the L-R upper body associations (i.e., hand movements) in the x coordinates regardless of age. Figure 6CF further depicts the association between body parts in tanglegram (R library “dendeextend” (https://talgalili.github.io/dendextend/)) in terms of “correct” or “incorrect” labels using manual codings. Regarding latency (Fig. 6C), the incorrect matrix for 2-year-olds is characterized where upper (R)-lower (R) parts are associated (red ellipse), while the other tanglegrams suggest that the upper (R) cluster was independently moving (yellow

rectangle). Regarding reaching (Fig. 6F), the incorrect matrix for 3-year-olds is characterized where the upper (R)-lower (R) parts are associated again (red ellipse), while the other tanglegrams suggest that the upper (R) and face clusters were associated in the x coordinates (yellow rectangle) (see Fig. S7A for the cross-correlation between the face and hand). If face movement is critical for precise touching (see Fig. 5D again), this could be related to their FB/FF control in the AR mark test (Shen et al., 2010). This might have emerged as the weight between the latency/reaching duration as a function of FB/FF control strategy between both age groups as follows.

## Feedforward/Feedback Control Weights Changes Among Ages

### *The Weight Between Feedback/Forward Strategy Changes Among Ages*

Thus far, the Bodytoypo task mainly generated participants' touch errors and the latency/reaching duration (Fig. 7B) as well as the association between body parts. These depend on the target location (see Fig. 5 and 6). This specifically varies the initial mark distance between the hand used and the target location when the mark appeared (Figure S6 for a descriptive correlation matrix among variables). Although categorical comparisons, such as 2- or 3-year-olds or correct/incorrect trials, have been effective in qualitatively depicting what was happening in the current task during the latency or reaching periods, the all-in-one statistics with a linear mixed model attempted to conclude these relationships continuously, using other covariates as mentioned above. Therefore, the best-fit model was explored using the R library ("lme4" (https://github.com/lme4/lme4/) and "lmerTest" (https://github.com/runehaubo/lmerTestR), where the initial inputs were latency duration, reaching duration, trial repetition, initial mark distance, and participants' age in months as fixed effects, as well as participants' ID and target labels as random effects; the touch error value was the output. This every-effect-model generated significant interaction effects (e.g., the interactions among reaching, age, repetition, and initial distance ($t(422.5) = 2.721$, $p =$

.0068)). The step function further identified the model with the minimum Akaike information criterion (AIC). This selected-effect-model still generated a significant interaction among latency/reaching, age, repetition, and initial distance ($t$(557. 0) = 2.30, $p$ = .0218 for latency; $t$(554. 0) = 2.38, $p$ = .0177 for reaching). Figure 7A presents the relationship between latency, reaching, and age in months in terms of predicting the touch errors (see Fig. S7B for other interactions when total RT was used instead of latency/reaching duration).

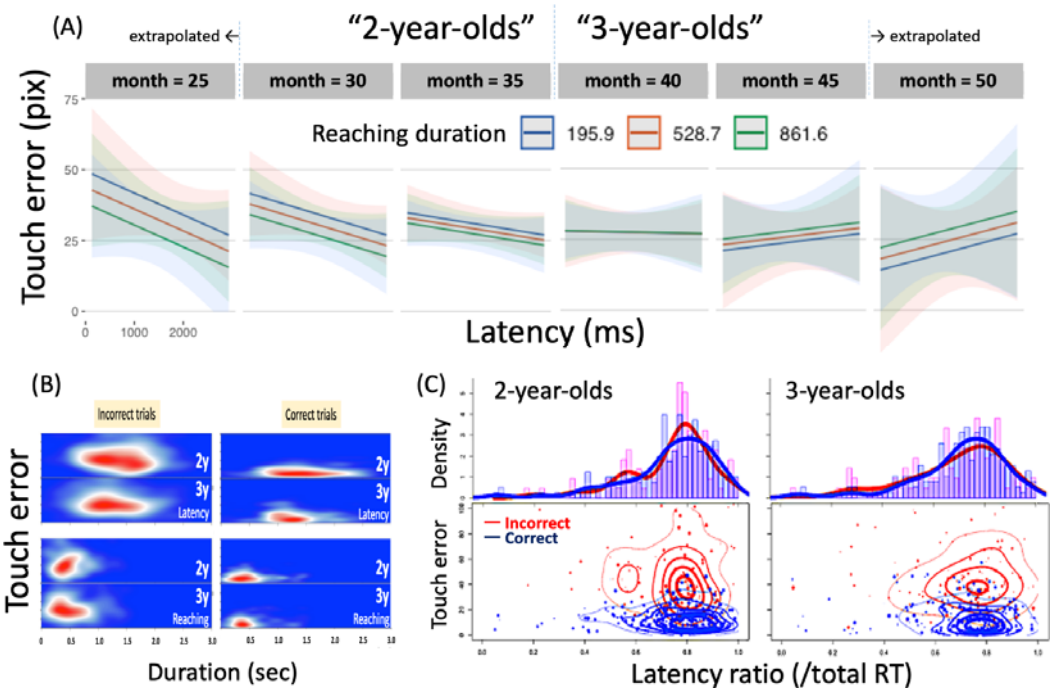


**Fig. 7.**

*Estimated Parameters for Predicting the Touch Error Regarding Age in Months*

*Note*. (A) A linear mixed model, fitted using the Akaike Information Criterion (AIC), to determine the most informative inputs. The model examined how latency, reaching duration, and age (in months) could predict touch errors. Interestingly, the slopes of these predictors gradually reversed as children got older. (B) Heatmaps to visualize the distribution of latency/reaching duration and touch errors across different ages and

feedback/forward weights.

For younger months, both latency and reaching should be longer for correct touch (minimized touch error). However, generally, there was a robust anti-correlation between latency and reaching duration (Fig. S6). For older months, however, these should be shorter for correct touch. This might imply that younger months rely on slower FB control when correct. Accordingly, if they administer a quick movement without visual feedback (i.e., ballistic exploration), they might make mistakes (see Fig. 7C; as shown below, the latency ratio did not vary for incorrect trials for 2-year-olds). Several months later, they no longer exhibited ballistic movements. The older participants may rely on FF control with a shorter latency/reaching when correct. Accordingly, if they exhibit longer reaching as FB control, they may mistouch. This finding may conflict with a conventional understanding of FB/FF weight with development (Bushnell, 1985). Given that visually guided reaching through a mirror (e.g., mirror drawing (Starch, 1910)) can confuse even an adult; more weight on FB control for the AR mark test is not the optimal strategy. Indeed, the trial repetition factor suggests that, for older months, the repetition increased their touch errors, presumably owing to habituation-driven exploration or playing with FB control (Fig. S7B, the upper panels).

Figure 7B shows the weight between latency/reaching duration in relation to the touch error and correct/incorrect labels. During a reach, the duration distribution peaked for 2-year-olds in incorrect trials and peaked for 3-year-olds in correct trials. When the relative duration of latency (the ratio for total RT) was summarized between 2- and 3-year-olds (Fig. 7C, a contour plot for visualization), the target-dependent variance of the latency ratio might be observed for correct trials with 2-year-olds, and incorrect trials with 3-year-olds, which could be evaluating their feedback weight. However, for the correct trials among 3-year-olds, the initial mark distance was positively correlated with the latency ratio (Fig. S6) and interacted with the total RT for touch errors (Fig. S7B, the lower panels). The hand was synchronized

with face movements in y-coordinates (Fig. S7A), which could be evaluated by their FF weight. The best strategy for the Bodytoypo task should be the optimal integration of FB/FF control depending on the target location (e.g., initial mark distance). The current results suggest a developmental weight shift from ballistic explorative movements (incorrect trials) to a FB-based slow control (correct trials) for 2-year-olds and from a FB-based slow control (incorrect trials) to a FF-based fast control (correct trials) for 3-year-olds. The current task could reveal the complex but optimal weight between the internal FF output, as well as the external FB input. We further re-considered visual observations over the original videos, according to the statistical results.

**Reaching Strategies and Evaluating Weights**

Automatic analysis revealed developmental changes in toddlers' reaching strategies: ballistic search, FB control, and FF control. Ballistic search refers to the participant reaching for the mark quickly, without aiming (Supplementary Video S1). FB control is a slower reaching strategy in which hand movements are adjusted while watching the video monitor or actual body part (Supplementary Video S2). Lastly, FF control is a strategy for reaching the target quickly while anticipating localization (Supplementary Video S3).

To examine the validity of this categorization, we tested whether the three reaching strategies found in the automatic analysis were classifiable to human observers: two coders reviewed the video and classified all trials into one of the three strategies. An agreement was calculated using 32.5% of the total trials. However, the agreement was low ($\kappa$ = .33), especially in the distinction between the ballistic exploration and FF, and between FF and FB; a difference of approximately 25% in judgment. This result suggests that the classification of reaching strategies is difficult for the human eye. Instead, we thought it reasonable to explain the switching of reaching strategies during the same trial (Supplementary Video S4) and the remaining ambiguity of reaching methods as an online

weighting of the internal model rather than as an exclusive choice of three strategies. However, the qualitative re-evaluation revealed that differences in the shape of the hand used for reaching, especially the use of the index finger (see Discussion), are valid indicators of developmental changes in reaching strategies (Fig. S3).

## Discussion

This study aimed to demonstrate the potential of the mark test to reveal toddlers' sensorimotor body representation. Focusing on the three properties of the mark test (self-oriented, visually guided, and spontaneous reaching), and through the analysis of the first touch of the Bodytoypo, we described body representation development from multiple perspectives.

Regarding semantic body representation, touch errors and vocabulary acquisition scores for each body part were not correlated. However, from the perspective of multimodal integration, touching body parts used in daily life, such as facial, joint, and movable parts, was relatively easy regardless of age. Furthermore, from the perspective of internal model acquisition, three types of reaching strategies could be assumed: ballistic search, FB control, and FF control. The strategies and weightings varied with age. These reaching strategies indicate a process of optimal integration of internal outputs and external inputs. This finding supports our previous observation regarding the rear search error in the classical mark test, which disappears among 2- to 3-year-olds (Imaizumi et al., 2021).

Evaluating body representations from the three perspectives may have created a new avenue for approaching the general question of self-recognition. Research on enhancing one's understanding of self-mirrored images based on one's own actions is scant. The self-image/mirror image is not merely an image of oneself but an object that can be controlled. This may elucidate the early development of a sense of body ownership and subjectivity alternative to the RHI task. Some possible future extensions include Bodytoypo, which

inserts a time delay or uses a self-image with arms unnaturally extended. Bodytoypo is also a task that participants can intuitively perform without verbal instructions, even young children.

The significance of the new technical approaches in our task is twofold: The first is the integration of new technologies with behavioral experiments. For real-time skeletal detection of the whole body, we used OpenPose, which detects the skeleton from 2D images and performs posture estimation using deep learning, allowing even small infants to detect the skeleton similar to adults (Ossmy & Adolph, 2020; Rocha & Addyman, 2022). Detecting motion trajectories without markers is significant for research with young children, because in another study using 3D markers (Kahrs et al., 2014), the attrition rate of participants was nearly twice that in this study due to marker removal (attrition rate: 31.8%). AR markers were displayed on each body part in near real-time using an advanced program. However, participants' touch judgments were made by manual control because automatic judgements based on 2D-image processing do not necessarily fit the touch subjectivity (even if the participants' arm just passed over the mark, the image processing would be touch judgement). By appropriately introducing manual judgments by the experimenter, we succeeded in making participants learn a rule that a touch to the virtual mark on the screen is a reward.

The second is automating the analysis by incorporating machine learning. In previous developmental studies, analyses were mainly based on human classification of actions in video images of children. However, a series of actions may involve multiple strategies; the first touches in this study were no exception. The significance of our analysis is to show changes in the weighting of reaching strategies, not classifications.

Therefore, this project is not simply a digitization of the mark test but a rich collaboration of engineering (online task processing), machine learning statistics (offline analysis of whole-body movements), and developmental psychology (theoretical consideration of crucial phenomena and conducting experiments).

The changes from ballistic exploratory movements to FB and/or FF control between the ages of two and three provide important insight into the acquisition of body representation, especially regarding the role of visual information in sensorimotor control. In contrast to ballistic exploratory movements, FB control requires online visual adjustment while referring to the learned self-body representation. This suggests that body part localization through ballistic exploratory movement weighs more on proprioceptive information, while FB/FF control should be an integrated process between proprioceptive and visual information (Sailer et al., 2005). Particularly in 3-year-olds, correct reaching showed more weight to FF than to FB control. Localization while observing the mirror image requires FB control, which is sometimes confusing even for adults. Index finger use, observed only during FF control, may also reflect appropriate motor prediction. This indicates the development of our sensorimotor integration with reliability among proprioceptive, visual, and motor predictions. The weights should always change depending on the circumstances or tasks. Our results suggest that 3-year-olds can adapt quickly as latency and reaching were rapid while utilizing motor prediction (FF weights for correct touch).

This process of taking in visual information has also been observed in studies of RHI and self-reaching. Children seem to prioritize visual information more than adults and, therefore, respond more to the visible hand than the felt hand (Cowie et al., 2018). Additionally, studies that placed vibrators on various body parts and analyzed the relevant responses show that reaching while looking at the target is not observed before six months (Somogyi et al., 2018). A certain amount of time and motor experience may be required for us to balance and smoothly incorporate visual information into our proprioceptive processing of self.

In addition to integrating vision and proprioception, young children gradually acquire the ability to control their bodies as objects or representations in a predictive manner. The classic mark test would remain useful in tandem with recent technical advantages.

## References

Amsterdam, B. (1972). Mirror self-image reactions before age two. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *5*(4), 297–305. https://onlinelibrary.wiley.com/doi/abs/10.1002/dev.420050403

Anderson, J. R., & Gallup, G. G., Jr. (2015). Mirror self-recognition: a review and critique of attempts to promote and engineer self-recognition in primates. *Primates; Journal of Primatology*, *56*(4), 317–326. https://doi.org/10.1007/s10329-015-0488-9

Bernstein, N. (1967). *The co-ordination and regulation of movements*. Pergamo Press, London.

Brownell, C. A., Nichols, S. R., Svetlova, M., Zerwas, S., & Ramani, G. (2010). The head bone's connected to the neck bone: when do toddlers represent their own body topography? *Child Development*, *81*(3), 797–810. https://doi.org/10.1111/j.1467-8624.2010.01434.x

Brownell, C. A., Zerwas, S., & Ramani, G. B. (2007). "So big": the development of body self-awareness in toddlers. *Child Development*, *78*(5), 1426–1440. https://doi.org/10.1111/j.1467-8624.2007.01075.x

Bushnell, E. W. (1985). The decline of visually guided reaching during infancy. *Infant Behavior & Development*, *8*(2), 139–155. https://doi.org/10.1016/S0163-6383(85)80002-3

Camões-Costa, V., Erjavec, M., & Horne, P. J. (2011). Comprehension and production of body part labels in 2- to 3-year-old children. *The British Journal of Developmental Psychology*, *29*(Pt 3), 552–571. https://doi.org/10.1348/026151010X523040

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2021). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(1), 172–186.

https://doi.org/10.1109/TPAMI.2019.2929257

Claxton, L. J., Keen, R., & McCarty, M. E. (2003). Evidence of motor planning in infant reaching behavior. *Psychological Science*, *14*(4), 354–356. https://doi.org/10.1111/1467-9280.24421

Corbetta, D., Thurman, S. L., Wiener, R. F., Guan, Y., & Williams, J. L. (2014). Mapping the feel of the arm with the sight of the object: on the embodied origins of infant reaching. *Frontiers in Psychology*, *5*, 576. https://doi.org/10.3389/fpsyg.2014.00576

Corbetta, D., Wiener, R. F., Thurman, S. L., & McMahon, E. (2018). The embodied origins of infant reaching: implications for the emergence of eye-hand coordination. *Kinesiology Review*, *7*(1), 10–17. https://doi.org/10.1123/kr.2017-0052

Cowie, D., Makin, T. R., & Bremner, A. J. (2013). Children's responses to the rubber-hand illusion reveal dissociable pathways in body representation. *Psychological Science*, *24*(5), 762–769. https://doi.org/10.1177/0956797612462902

Cowie, D., McKenna, A., Bremner, A. J., & Aspell, J. E. (2018). The development of bodily self-consciousness: changing responses to the full body illusion in childhood. *Developmental Science*, *2*(3 ), e12557. https://doi.org/10.1111/desc.12557

Cowie, D., Sterling, S., & Bremner, A. J. (2016). The development of multisensory body representation and awareness continues to 10 years of age: evidence from the rubber hand illusion. *Journal of Experimental Child Psychology*, *142*, 230–238. https://doi.org/10.1016/j.jecp.2015.10.003

de Waal, F. B. M. (2019). Fish, mirrors, and a gradualist perspective on self-awareness [Review of *Fish, mirrors, and a gradualist perspective on self-awareness*]. *PLoS Biology*, *17*(2), e3000112. https://doi.org/10.1371/journal.pbio.3000112

Gallup, G. G., Jr. (1970). Chimpanzees: self-recognition. *Science*, *167*(3914), 86–87. https://doi.org/10.1126/science.167.3914.86

Gottwald, J. M., Bird, L.-A., Keenaghan, S., Diamond, C., Zampieri, E., Tosodduk, H.,

Bremner, A. J., & Cowie, D. (2021). The developing bodily self: How posture constrains body representation in childhood. *Child Development*, *92*(1), 351–366. https://doi.org/10.1111/cdev.13425

Hofsten, C. von, & von Hofsten, C. (1984). Developmental changes in the organization of prereaching movements. *Developmental Psychology*, *20*(3), 378–388. https://doi.org/10.1037/0012-1649.20.3.378

Imaizumi, S., Asai, T., & Miyazaki, M. (2021). Cross-referenced body and action for the unified self: empirical, developmental, and clinical perspectives. In Y. Ataria, S. Tanaka, & S. Gallagher (Eds.), *Body schema and body image: New directions* (pp. 194–209). Oxford University Press. https://books.google.co.jp/books?hl=ja&lr=lang_ja|lang_en&id=ObEzEAAAQBAJ&oi=fnd&pg=PA194&dq=Imaizumi+Asai+Miyazaki&ots=c3GL0r1h_l&sig=CyixjqQUBFrpy22dSu7nCh5nFpw

Kahrs, B. A., Jung, W. P., & Lockman, J. J. (2014). When does tool use become distinctively human? hammering in young children. *Child Development*, *85*(3), 1050–1061. https://doi.org/10.1111/cdev.12179

Lewis, M., & Brooks-Gunn, J. (1979). *Social cognition and the acquisition of self*. Springer New York https://doi.org/10.1007/978-1-4684-3566-5

Miyazaki, M., Asai, T., & Mugitani, R. (2019). Touching! an augmented reality system for unveiling face topography in very young children. *Frontiers in Human Neuroscience*, *13*. https://doi.org/10.3389/fnhum.2019.00189

Miyazaki, M., & Hiraki, K. (2017). *Does rear-search error in the mark test indicate a uniqueness of body-representation in young children?* The sixth annual Budapest CEU Conference on Cognitive Development, Budapest, Hungary. https://www.bcccd.org/down/bcccd_final_program_2017.pdf

Ossmy, O., & Adolph, K. E. (2020). Real-time assembly of coordination patterns in

human infants. *Current Biology: CB*, *30*(23), 4553–4562.e4. https://doi.org/10.1016/j.cub.2020.08.073

Rocha, S., & Addyman, C. (2022). Assessing sensorimotor synchronisation in toddlers using the Lookit online experiment platform and automated movement extraction. *Frontiers in Psychology*, *13*, 897230. https://doi.org/10.3389/fpsyg.2022.897230

Rochat, P., & Zahavi, D. (2011). The uncanny mirror: a re-framing of mirror self-experience. *Consciousness and Cognition*, *20*(2), 204–213. https://doi.org/10.1016/j.concog.2010.06.007

Sailer, U., Randall Flanagan, J., & Johansson, R. S. (2005). Eye–hand coordination during learning of a novel visuomotor task. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *25*(39), 8833–8842. https://doi.org/10.1523/JNEUROSCI.2658-05.2005

Schwoebel, J., & Coslett, H. B. (2005). Evidence for multiple, distinct representations of the human body. *Journal of Cognitive Neuroscience*, *17*(4), 543–553. https://doi.org/10.1162/0898929053467587

Scott Kelso, J. A. (2014). *Human motor behavior: an introduction* (J. A. S. Kelso (ed.)). Psychology Press. https://doi.org/10.4324/9781315802794

Shen, H., Baker, T. J., Candy, T. R., Yu, C., & Smith, L. B. (2010). Using the head to stabilize action: reaching by young children. *2010 IEEE 9th International Conference on Development and Learning*, 108–113. https://doi.org/10.1109/DEVLRN.2010.5578856

Somogyi, E., Jacquey, L., Heed, T., Hoffmann, M., Lockman, J. J., Granjon, L., Fagard, J., & O'Regan, J. K. (2018). Which limb is it? responses to vibrotactile stimulation in early infancy. *The British Journal of Developmental Psychology*, *36*(3), 384–401. https://doi.org/10.1111/bjdp.12224

Starch, D. (1910). A demonstration of the trial and error method of learning.

*Psychological Bulletin*, *7*(1), 20. https://psycnet.apa.org/journals/bul/7/1/20/

Suddendorf, T., & Butler, D. L. (2013). The nature of visual self-recognition. *Trends in Cognitive Sciences*, *17*(3), 121–127. https://doi.org/10.1016/j.tics.2013.01.004

Waugh, W., & Brownell, C. (2015). Development of body part vocabulary in toddlers in relation to self-understanding. *Early Child Development and Care*, *185*(7), 1166–1179. https://doi.org/10.1080/03004430.2014.983915

Witt, A., Cermak, S., & Coster, W. (1990). Body part identification in 1- to 2-year-old children. *The American Journal of Occupational Therapy*, *44*(2), 147–153. https://doi.org/10.5014/ajot.44.2.147