Received: 9 March 2022 | Accepted: 22 February 2023

DOI: 10.1111/bjop.12644

ARTICLE

the british psychological society
promoting excellence in psychology

Paradoxical decrease of imitation performance with age in children

Giovanni Ottoboni[1] | Alessio Toraldo[2,3] | Riccardo Proietti[1] | Angelo Cangelosi[4] | Alessia Tessari[1,5]

[1]Department of Psychology 'Renzo Canestrari', University of Bologna, Bologna, Italy

[2]Department of Brain and Behavioural Sciences, University of Pavia, Pavia, Italy

[3]Milan Centre for Neuroscience (NeuroMI), Milan, Italy

[4]Department of Computer Science, University of Manchester, Manchester, UK

[5]Alma Mater Research Institute for Human-Centered Artificial Intelligence, Bologna, Italy

Correspondence

Alessia Tessari, Department of Psychology, University of Bologna, viale Berti Pichat, 5, 40127, Bologna, Italy.
Email: alessia.tessari@unibo.it

Alessio Toraldo, Department of Brain and Behavioural Sciences, University of Pavia, Piazza Botta 11, 27100, Pavia, Italy.
Email: alessio.toraldo@unipv.it

Abstract

Imitation development was studied in a cross-sectional design involving 174 primary-school children (aged 6–10), focusing on the effect of actions' complexity and error analysis to infer the underlying cognitive processes. Participants had to imitate the model's actions as if they were in front of a mirror ('specularly'). Complexity varied across three levels: movements of a single limb; arm and leg of the same body side; or arm and leg of opposite body sides. While the overall error rate decreased with age, this was not true of all error categories. The rate of 'side' errors (using a limb of the wrong body side) paradoxically increased with age (from 9 years). However, with increasing age, the error rate also became less sensitive to the complexity of the action. This pattern is consistent with the hypothesis that older children have the working memory (WM) resources and the body knowledge necessary to imitate 'anatomically', which leads to additional side errors. Younger children might be paradoxically free from such interference because their WM and/or body knowledge are insufficient for anatomical imitation. Yet, their limited WM resources would prevent them from successfully managing the conflict between spatial codes involved in complex actions (e.g. moving the left arm and the right leg). We also found evidence that action side and content might be stored in separate short-term memory (STM) systems: increasing the number of sides to be encoded only affected side retrieval, but not content retrieval; symmetrically, increasing the content (number of movements) of the action only affected content retrieval, but not side retrieval. In conclusion, results suggest that anatomical imitation might

Giovanni Ottoboni and Alessio Toraldo contributed equally to this work.

interfere with specular imitation at age 9 and that STM storages for side and content of actions are separate.

KEYWORDS
body knowledge, children, development, double dissociation, imitation, meaningless action, movement complexity, working memory

BACKGROUND

Imitating others' actions is crucial in acquiring new knowledge through social learning in infants and children. Human beings imitate their fellow humans to acquire several new abilities by following their example, and such ability seems to originate early in life (Carpenter et al., 1998; Heyes, 2001; Meltzoff, 1996). Imitation is a key learning mechanism in humans, which allows for rapid transfer of knowledge among individuals (Tomasello, 2003): observing someone's actions allows humans to learn faster and more easily to relate with objects and the environment over time, especially during early childhood (Meltzoff et al., 2009).

Importantly, the term 'imitation' refers to the ability to learn to perform an action and the ways to achieve it by reproducing both observed goals and gestures (Longo et al., 2008; Lyons et al., 2011; Want & Harris, 2002; Zentall, 2012). Thus, for *proper* imitation to occur, both the goals and the means to achieve them must be reproduced. This contrasts with 'emulation', where only the goals are reproduced, and individuals learn about characteristics of objects and the environment without necessarily learning actions themselves. Importantly, these two processes seem to follow different developmental trajectories (Jones, 2009; McGuigan & Whiten, 2009; Want & Harris, 2002).

Imitation is recognized to rely on complex cognitive mechanisms involving observation, planning and action control. For instance, the *goal-directed theory* (GOADI; Bekkering et al., 2000; Wohlschläger et al., 2003) assumes that individuals first decompose the observed action into its constituent motor patterns and then recombine them to reconstruct the original action. This decomposition/reconstruction process is guided by an internal representation of the action in terms of goals rather than motor segments. More precisely, a *hierarchy* of goals, ordered by relevance, is obtained, and the individual will reproduce as many goals (starting from the top of the hierarchy) as their working memory (WM; Baddeley, 2010) capacity can handle. Therefore, if an action is very complex or new, only the main goals (higher in the hierarchy) will be reproduced due to WM limits. GOADI assumes that the processes involved in imitation are exactly the same in both children and adults, with the only difference concerning WM resources (Wohlschläger et al., 2003). GOADI also states that a goal directly elicits the motor program most strongly associated with the seen action, highlighting the role of (procedural) long-term memory. The *dual-route model* (Cubelli et al., 2000; Rothi et al., 1991; Rumiati & Tessari, 2002) elaborates on two independent processes of imitation: a direct route, used for imitating new, meaningless actions, which, however, can also reproduce known, meaningful ones when necessary, and a 'lexical-semantic' (indirect) route, which capitalizes on a long-term memory storage of known actions (the 'lexicon') and can only be used to imitate this kind of material (Rothi et al., 1991). Some authors also focused on the role of WM in the processing of the direct route, as the action's subunits need to be stored until they are reassembled in the motor output (Buxbaum & Randerath, 2018; Cubelli et al., 2000; Rumiati & Tessari, 2002; Tessari & Cubelli, 2014; Toraldo et al., 2001). In this view, the role of WM is very similar to the one proposed by GOADI, as WM limits can bound imitative performance (the poorer the WM capacity, the worse the imitation performance; see, for example, results from double-task WM interference: Rumiati & Tessari, 2002; Tessari & Rumiati, 2002). Moreover, WM allows the two routes to interact in a shared working space to learn new actions (Tessari et al., 2006). Even though both routes are activated simultaneously (much like in other famous dual-route models, e.g. Coltheart et al., 2001), the specific demands of an imitation task can modulate the emphasis on the information coming from one route or the other (Cubelli et al., 2006; Ottoboni et al., 2018; Tessari & Cubelli, 2014; Tessari & Rumiati, 2004; Tessari et al., 2021, for a critical discussion) and the available cognitive resources play an essential role (Tessari et al., 2006).

Error analysis

Both the GOADI and the dual-route model relied not only on quantitative measures as accuracy rates but also on finer, qualitative analysis of the specific error types. This 'error analysis' approach has proved very fruitful in (neuro)psychology in general (e.g. McCloskey, 2003; Newcombe & Marshall, 1973; Toraldo & Shallice, 2004) and provided critical insight also in the field of action imitation. Based on studies of the dual-route theory, Tessari and Rumiati (2004) assumed that a *homogeneous* list of meaningful actions would have led to the use of the lexical-semantic route, which is specialized for that type of stimulus. By contrast, a *mixed* list of meaningful and meaningless actions would have activated the direct, non-semantic route, because this can process both stimulus types. Beyond accuracy measures, a convincing proof that the semantic route had indeed been used in the former condition was that the frequency of semantic errors (prototypicalization, body part as a tool, visuosemantic) was higher among the meaningful items of the homogeneous lists than among the same items of the mixed lists. Similar results were recorded when the effect of list composition was investigated in brain-damaged patients (Tessari et al., 2007): the two patients with an impaired semantic route showed a higher frequency of semantic errors when imitating meaningful actions in a homogeneous list than in a mixed list. Hence, the authors capitalized on evidence obtained from error analysis.

The GOADI theory itself arose from the analysis of imitation errors in a simple hand-to-ear task (Bekkering et al., 2000). In this task, a model touches one or (simultaneously) both of their own ears with their hand(s), and the subject has to imitate the action(s). The reaching movements can be ipsilateral or contralateral (e.g. right hand to, respectively, the right or the left ear). Preschool children (3–5 years) often (40%) imitated a contralateral movement with an ipsilateral one to the correct target. Thus, for instance, they touched their left (correct) ear with their left (wrong) hand instead of their right (correct) hand. By analysing the quality of the error types, the authors inferred that the primary goal—the 'correct' ear to be reached—dominated over the secondary goals, like the hand to be used to perform the movement, and assumed that such goal selection was necessary because of the WM limits characterizing that young population. Indeed, when the target ear was kept constant in a second experiment and thus did not need to be maintained in WM, the children could focus on the movement, and wrong ipsilateral reaching movements almost disappeared. Once again, error-type analysis was critical and allowed the authors to develop their theory on the hierarchy of goals.

Imitation in children

The study of imitation in children has long been of interest in the psychological literature (e.g. Hurley & Chater, 2005; Meltzoff, 1988; Want & Harris, 2002; Wohlschläger et al., 2003). Developmental studies on imitation have generally focused on infants and preschool children, to assess the presence of imitative mechanisms suggestive of the acquisition of specific cognitive abilities (Dickerson et al., 2012; Huang & Charman, 2005; Nielsen, 2006). The question of *when* children and infants begin to imitate is still a matter of debate. Some authors propose that imitation is innate (Meltzoff & Moore, 1997; Nagy et al., 2005; Simpson et al., 2014), whereas others propose that it develops during the first years of life (Jones, 2007; Ray & Heyes, 2011). Some studies suggest a later development for imitation: children would become good imitators by age 3 (Horner and Whiten, 2005; McGuigan et al., 2007; Piaget, 1962). However, up to age 3, imitation is effortful, and its automatization would occur later, between ages 3 and 7 (O'Sullivan et al., 2018).

Imitation in children seems to depend critically on social context factors. This is clearly evident in research on motor synchronization. Interpersonal rhythmic coordination can occur both intentionally and unintentionally (relying on visual information: e.g. Richardson et al., 2007; Schmidt & Richardson, 2008; Temprado & Laurent, 2004) and has been shown to depend on social variables (e.g. Tunçgenç & Cohen, 2016). Howard et al. (2021) demonstrated that the social context influences synchrony between participants in both adults and children. The effect is particularly strong in the latter group when an explicit request is made by the experimenter.

In most experimental paradigms dealing with imitation, the social context is essentially driven by the *model*—the actor whose action is to be imitated. Three main factors seem to contribute to the role of the model: perceived authoritativeness, familiarity with the action, and motivation. Previous work suggested that an adult model acquires a role of behavioural guidance on children's imitation because adults are perceived as figures of authority. When *new* (unfamiliar) actions are shown, children tend to rely more on adults than on peers, as they tend to acquire information from adults for unfamiliar domains but equally often from peers and adults for familiar domains (e.g. Taylor et al., 1991). Children infer the expertise and the reliability of the model based on the inferred expertise on specific domains (Chow et al., 2008; Jaswal & Malone, 2007; Jaswal & Neely, 2006; Ottoboni et al., 2013), and adults are considered, in general, more reliable in light of their role for social learning (Laland, 2004). This notion was confirmed by Zmyj and Seehagen (2013), who reviewed a large corpus of literature and concluded that children prefer to imitate peers on familiar actions and adults on new ones (see, e.g. McGuigan et al., 2011; Wood et al., 2012; Rakoczy et al., 2010; these authors reported a bias in favour of adult models by children who imitated novel, unusual, irrational or irrelevant actions). As to motivational factors, the most important one in children's imitation is likely to be 'the need to belong' (e.g. Baumeister & Leary, 1995; Fiske, 2010; Over, 2016). Imitation is also a means of forming and maintaining relationships with others and is sometimes called 'social' imitation (i.e. Over & Carpenter, 2013). In particular, children imitate accurately when they want to join someone or a group (Over & Carpenter, 2009), they tend to help others more often after an adult has mimicked them (Carpenter et al., 2013), and they infer important information on relationships between third parties by observing their imitative behaviour (Powell & Spelke, 2018). The model a child decides to imitate plays a crucial role in social pressure (e.g. Haun & Tomasello, 2011) or in achieving social goals such as making friends. In imitative behaviour, therefore, social motivation is critical (Hoehl et al., 2019) and can be considered a form of 'social glue' (Lakin et al., 2003).

Despite this rich line of research, several aspects of children's imitation have been relatively neglected; the purpose of the present work was to fill these gaps.

Aims of the study

The relatively overlooked aspects we wished to investigate were: (i) the error profiles, over and above simple accuracy scores; (ii) the role of action complexity; (iii) the comparison between human and non-human models and (iv) the development of imitation performance through school age (6–10 years). As the reader will see, this approach proved fruitful and led to unexpected results, which in turn, should constrain theories about the cognitive processes involved in children's imitation.

Error analysis and action complexity

Imitation in children has been typically investigated with an eye on overall accuracy (with a few notable exceptions, including the above-mentioned Bekkering et al.'s, 2000 GOADI study, and some earlier work, e.g. Benton, 1959; Gordon, 1923; Schofield, 1976; Wapner & Cirillo, 1968). Children are known to be able to appropriately parse complex visual stimuli (e.g. Bauer, 1992) and to imitate complex, sequentially structured actions (e.g. Whiten et al., 2009). However, relatively little is known about the error profiles with complex actions. We aimed at exploiting the heuristic potential of error analysis to gain new insight into the cognitive processes involved in this stimulus type.

In the present study, the level of complexity of the action was explicitly manipulated in order to vary the cognitive load in participants. We included single-limb movements (the simplest ones), two-limb 'homolateral' movements (involving the arm and the leg of the same body side) and two-limb 'hetero-lateral' movements (the most complex ones, involving the arm of one side and the leg of the opposite side). Only new, *meaningless* actions were included to prevent children from relying on previously acquired knowledge. Indeed, a semantic strategy of imitation would have likely swamped the effects of differences in action complexity.

Human versus non-human models

The lack of meaning in the actions also allowed our participants to focus more on the model, hence possibly enhancing the effects of this social variable. Beyond the well-studied adult and child models, we also introduced a humanoid child robot (iCub). Robots have proved effective in promoting autistic children's imitation (e.g. Conti et al., 2015; Duquette et al., 2008; Zheng et al., 2014), but data from typically developing children are scarce (e.g. Srinivasan et al., 2013).

Age

Through a cross-sectional approach, this study investigated how imitation develops in first-grade school children (6–10 years). We focused on this age range because most studies (e.g. Horner & Whiten, 2005; McGuigan et al., 2007) have concentrated on early childhood (with the general agreement that humans are good imitators by age 3). Late childhood (first school years) has been relatively neglected; however, in this developmental phase, WM, an essential component in imitation according to both GOADI and dual-routes theories, is not fully developed yet, as it mainly develops during adolescence (Burnett Heyes et al., 2004; Gathercole et al., 2004; Heyes et al., 2016; Isbell et al., 2015; Ullman et al., 2014). Some studies did investigate the primary-school age group (e.g. O'Sullivan et al., 2018), but they did not study the roles of action complexity and model.

Predictions

While this study was exploratory in essence, we did have some predictions. As to the model variable, we predicted a higher performance with an adult model than a child or robot model, given that meaningless actions should make children rely more on the authoritative adult figure (Taylor et al., 1991). Also, motivational factors, like the need for affiliation (Over & Carpenter, 2012, 2013) justified the same prediction: given that the only person present in the lab during the experiment would have been an adult (the experimenter), the children might have implicitly tried to please this person by better imitating the adult model. Moreover, we expected imitation performance to improve with age as a general effect of the maturation of the children's cognitive system and, particularly, as an effect of the increase in WM (Gathercole, 1999; Isaacs & Vargha-Khadem, 1989; Wilson et al., 1987), which has been assumed to be a critical factor in imitation (Wohlschläger et al., 2003).

METHOD

Participants

One hundred and seventy-four children (88 females; mean age = 8.25 years, *SD* = 1.34, range 6–10, with only two 11 year olds) participated in the experiment. They were enrolled in primary schools in the city of Bologna, after the project had been presented to school Principals and families. All parents gave their consent to participation. All children reported normal or corrected-to-normal vision, were naïve to the

TABLE 1 Distribution of participants across experimental conditions (Model), Gender and Age (years).

Model	Participants' gender	Age 6	7	8	9	10–11
Adult	F	4	5	11	5	2
	M	4	4	6	6	10
Child	F	6	4	6	8	8
	M	4	1	9	11	5
Robot	F	2	6	9	5	7
	M	5	4	6	6	5

Note: Only two 11-year-olds participated in the study (both F, one with Adult, one with Robot model).

purpose of the experiment and were typically developing individuals (i.e. no cognitive delays, specific cognitive impairments, ADHD or specific learning disabilities). Each child was randomly assigned to one of the three model conditions (Table 1).

Ethics statement

Ethical approval was granted by the Bioethics Committee of the University of Bologna on May 7th, 2013, following the Declaration of Helsinki, and parents/tutors provided written informed consent for the children to participate in the study.

Stimuli

We used 12 meaningless actions (see Appendix C for description). Four actions involved only one limb (arm or leg; 'Single'-effector condition), four actions involved one arm and one leg belonging to the same side of the body ('Homolateral' condition), and four actions involved one arm and one leg belonging to two different sides of the body ('Heterolateral' condition; the movements were matched to those of the Homolateral condition, see Appendix C). Homolateral and Heterolateral actions will sometimes be collectively referred to as the 'Double'-effector conditions. Each action was performed by three different models: a 10-year-old male child, a 22-year-old man and an iCub robot (Figure 1), and videos were recorded of these performances (the iCub videos were created using the humanoid platform simulator iCubSim, Tikhanoff et al., 2008). The 36 videos were flipped along the horizontal axis to generate the left–right mirror image of each action, thus leading to 72 videos overall.

Procedure

Each child was shown only the 24 actions performed by one model. A laptop (Pentium III, 512 Mb), running E-Prime 1.1 software, SP3 (Schneider et al., 2012), controlled stimulus presentation and was connected to a keyboard and a DELL 2300 projector. The latter projected the action videos on a white wall 3 m from the children. Each child stood in front of a small table, on top of which lay the keyboard. Each trial began with the child keeping the keyboard's spacebar pressed while observing the video projected on the wall, which lasted for 5 s. At the end of the video, a go-signal (a red circle) was projected and informed the participants that they could release the spacebar and imitate the action. Children were explicitly required to imitate the actions specularly (i.e. as if they were in front of a mirror) and to pay attention to all of the action's features (e.g. the involved body parts, the exact movements and trajectories). We chose specular imitation as it is more natural in children (Brass et al., 2000). When the participants

FIGURE 1 Three examples of stimulus actions are shown. (a) The adult model performs a Single action. (b) The child model performs a Homolateral action. (c) The iCub robot model performs a Heterolateral action.

were satisfied that they had completed the required action, they pressed the spacebar again, and the next trial began. Keeping the spacebar pressed during the observation of the action video forced the children to reproduce the action after the video had ended, thus avoiding concurrent imitation. Children were encouraged and supported after each trial, irrespective of the quality of their imitative performance, to keep their attention constant and to prevent their motivation from dropping in case of mistakes. Imitation was video-recorded and later scored by two independent raters (Cohen's Kappa = .86). These reported and extracted all the error types found in the videos, and later discussed such categories with authors GO and Ate. Based on this work, and on error classifications that had already been reported in the literature, 10 different error types emerged, which are listed below.

- *Distal*: the hand (and/or the foot) is (are) moved incorrectly or not moved at all, while the involved limb(s) is (are) moved correctly (e.g. Carmo & Rumiati, 2009; Tessari et al., 2007; Tessari & Rumiati, 2004).
- *Proximal*: the movement(s) of the limb(s) is (are) correct or recognizable, but the angle(s) of the movement(s) with respect to the trunk is (are) wrong (e.g. Carmo & Rumiati, 2009; Tessari et al., 2007; Tessari & Rumiati, 2004).
- *Final Position*: wrong final position(s) of the involved limb(s) (Carmo & Rumiati, 2009).
- *Perseveration*: the performed action is completely different from the target one, but is similar to another previously observed action (global perseveration), or the performed action is partially correct and partially ascribable to another, previously observed action (partial perseveration). See (e.g. Carmo & Rumiati, 2009; Tessari & Rumiati, 2004; Tessari et al., 2007).
- *Omission*: one or more necessary steps of the target action are not carried out (e.g. Carmo & Rumiati, 2009; Tessari et al., 2007; Tessari & Rumiati, 2004).
- *Sequence*: movements of arm and leg that are synchronous in the target action are performed asynchronically, or temporally distinct components of a limb's movement are performed synchronically.
- *Side*: imitation is anatomical and not mirror-reversed for the arm, leg or both (Bekkering et al., 2000; Schofield, 1976; Wohlschläger et al., 2003).
- *Conduite d'approche*: the participant attempts to correct their mistakes through spontaneous corrections, with actions getting closer and closer to the target (Carmo & Rumiati, 2009; Della Sala et al., 2006; Howard et al., 2019).

- *Arm/Leg*: an observed arm movement is carried out with the leg or vice versa.
- *Unclassifiable*: the child produces an action that is completely different from the target one and also from any previously observed action (e.g. Carmo & Rumiati, 2009; Tessari et al., 2007; Tessari & Rumiati, 2004).

On Double-effector trials, an error of the Distal, Proximal, Final Position or Side type was diagnosed if it regarded at least one of the two limbs involved. By their very nature, Arm/Leg errors only applied to the Double-effector conditions. Imitation of an action (trial) was scored 'inaccurate' if at least one of the listed error categories applied.

Measures

The basic index we analysed was the plain error rate (referred to as 'raw ER' or 'per-trial ER' or simply 'ER', without specification), which indicates the relative frequency of trials with a given error type. A second index was introduced that helped us interpreting the frequently observed disadvantage of Double-effector with respect to Single-effector actions. Indeed, such a disadvantage is *per se* ambiguous: there are two main explanations for it, the 'mutual perturbation' and the 'independent processing' hypotheses. The 'mutual perturbation' hypothesis assumes that in Double-effector trials, the processing concerning a limb perturbs the processing concerning the other limb, and vice versa, with the effect of increasing the error rates of both; by contrast, the 'independent processing' hypothesis assumes that the two processes do not influence each other. Counterintuitively, also the latter predicts a higher ER for Double-effector trials. Indeed, albeit processed independently, two limbs are two 'opportunities' to make an error and since one error is enough to classify a Double-effector trial as incorrect, the mere fact that two limbs need processing will increase the ER (see Appendix B). For example, suppose that a child is required to copy either a single digit, 6, or two digits, 7–9. Even though the processing of 7 has no influence whatsoever on the (subsequent) processing of 9, the probability of making at least one mistake in the 7–9 pair is higher than the probability of making a mistake on the isolated 6.

So, both the 'mutual perturbation' and the 'independent processing' hypotheses are compatible with a disadvantage of Double-effector trials in terms of raw ER. However, another index can disentangle between those interpretations: the *per-limb* error rate ('per-limb ER'). If the processes regarding the two limbs were independent, the per-limb ER would come out identical for both Double- and Single-effector trials, while if there were mutual perturbations, Double-effector trials would still show a disadvantage. Thus, whenever the raw ER showed a disadvantage for Double- with respect to Single-effector conditions, we also analysed per-limb ERs. Mathematically, the per-limb ER corresponds to the relative frequency of limbs that were moved in an erroneous way, out of the overall number of limbs that had to be moved (Appendix B reports all details on how the index was empirically estimated). Since the difference in per-limb ER between the Double- and the Single-effector conditions informs us about the perturbation that the processing concerning a limb undergoes when another limb is 'added' to the action, we called this difference the 'mutual perturbation effect'.

Statistical analyses

Generalized Mixed Linear Models (GMLM) with Binomial distribution and logit link were used to analyse error rates (ERs). The models included the full factorial design Condition (Single effector vs. two Homolateral effectors vs. two Heterolateral effectors) × Model (Adult vs. Child vs. Robot) × Age (as a categorical variable). Given that Condition was the only within-subjects factor, the random design contained the by-subject random intercept and the Condition random slope. When Age effects were significant and theoretically interesting, we tried to understand at what age the ER 'flattens' or stabilizes. Indeed, the expected pattern is that ER decreases with age, till it reaches some minimum value and stops decreasing.

We could estimate the age at which such a minimum is (essentially) achieved using non-linear regression. However, surprisingly, ER increases with age in some cases, so here we estimated the age at which the ER curve begins to rise from a minimum, using the same form of non-linear regression. All mathematical details are reported in Appendix A.

Given that we analysed the overall ER, as well as nine specific ERs (those related to the ten error categories, minus one: 'unclassifiable' errors were not analysed), a Bonferroni correction was applied in the latter set: effects whose $p < .05/9 = .0055$ were considered to be significant.

Per-limb ERs could not be analysed by means of GMLM, because each estimate (see Appendix B) required aggregating data from more than one trial. Thus we applied General Equation Estimation (GEE), which allowed us to properly model the effect of a repeated-measure variable (Condition) on a variable characterized by very irregular (Tweedie) shapes.

Data and codes are provided on https://tinyurl.com/alessiotoraldo or https://osf.io/daewb/?view_only=bb85db7595a54082bf6900d1aef5064e.

RESULTS

Table 2 reports the absolute and relative frequencies of the various error categories. Table 3 reports the results from the GMLM analyses.

Overall error rate

The GMLM analysis detected a significant effect of Age ($F(4, 4131) = 2.609, p = .034$): ER showed some moderate drop with Age (Figure 2a). An exponential decay model estimated the stabilization age to be 8.97 ± 3.24 years (Appendix A). Thus, performance flattened at about age 9, but this parameter had a very unstable estimate (a standard error over 3 years). Looking at Figure 2a, such instability is likely related to an unexpected 'inversion' of the improving trend between ages 8 and 9, which will become clear from later error-specific analyses.

Condition yielded a significant effect ($F(2, 4131) = 32.241, p < .001$), but with a strong Condition × Model interaction ($F(4, 4131) = 7.631, p < .001$) shown in Figure 3a. Indeed, Single-effector actions

TABLE 2 Classification of errors.

	Raw error count	% out of all errors (*N* = 2024)	% out of all trials (*N* = 4176)
Distal	220	10.87	5.27
Proximal	486	24.01	11.64
Final position	218	10.77	5.22
Perseveration	158	7.81	3.78
Omission	50	2.47	1.20
Sequence	217	10.72	5.20
Side	604	29.84	14.46
Conduite d'approche	118	5.83	2.83
Arm/Leg	186	9.19	4.45
Unclassifiable	50	2.47	1.20
Double classifications[a]	271	13.39	6.49
Triple classifications[a]	6	0.30	0.14
Overall	2024	100.00	48.47

[a]Given the presence of double and triple classifications, single categories' error counts do not sum up at 2024, which is the real overall number of errors. To obtain 2024, one must subtract the number of double classifications and twice the number of triple classifications.

TABLE 3 Results from generalized mixed linear models (GMLM).

Predictor (df)	Overall ER (4131)	Distal (4131)	Proximal (4147)	Final (4131)	Persev. (4131)	Omission (4131)	Sequence (4131)	Side (4147)	Conduite d'appr. (4131)	Arm/Leg (2754)[a]
Age (4)	**2.609, .034**	1.003, .405	**11.566, <.001**	0.534, .711	0.48, .751	0.121, .975	2.867, .022	**9.864, <.001**	1.409, .228	0.674, .61
Condition (2)[a]	**34.241, <.001**	**76.627, <.001**	**40.543, <.001**	**14.321, <.001**	0.213, .808	0.42, .657	4.931, .007	**58.28, <.001**	3.314, .036	3.444, .064
Model (2)	0.463, .629	0.699, .497	0.907, .404	4.412, .012	1.249, .287	0.5, .607	1.996, .136	4.928, .007	1.636, .195	0.268, .765
AxC (8)[a]	1.115, .35	1.53, .141	1.435, .176	0.437, .9	1.33, .223	0.031, 1	1.097, .362	2.255, .021	0.358, .943	0.3, .878
AxM (8)	1.781, .076	0.098, .999	1.85, .063	0.999, .434	0.379, .932	0.237, .984	0.443, .895	1.118, .347	0.335, .953	0.89, .524
CxM (4)[a]	**7.631, <.001**	0.901, .462	0.927, .447	0.645, .63	1.001, .406	0.182, .948	**6.448, <.001**	1.507, .197	0.124, .974	0.227, .797
AxCxM (16)[a]	0.896, .574	0.184, 1		0.397, .984	0.403, .982	0.116, 1	0.544, .925		0.137, 1	0.351, .946

Note: Each cell reports *F* and *p*-values, which are in **bold** when significant (Bonferroni correction applied to specific error categories). Different rows report effects by different predictors. Interactions contain abbreviated variables names, A, Age (analysed as a categorical variable), C, Condition, M, Model. Different columns report different error categories (dependent variables). Values in brackets indicate degrees of freedom.

Abbreviation: ER, error rate.

[a]Arm/Leg confusion errors: Single-effector trials were omitted, as no Arm/Leg confusion errors could have occurred in this case; thus, the numerator's degrees of freedom were 1 for Condition, 4 for AxC, 2 for CxM, 8 for AxCxM. Empty cells: triple interactions were omitted due to convergence failures.

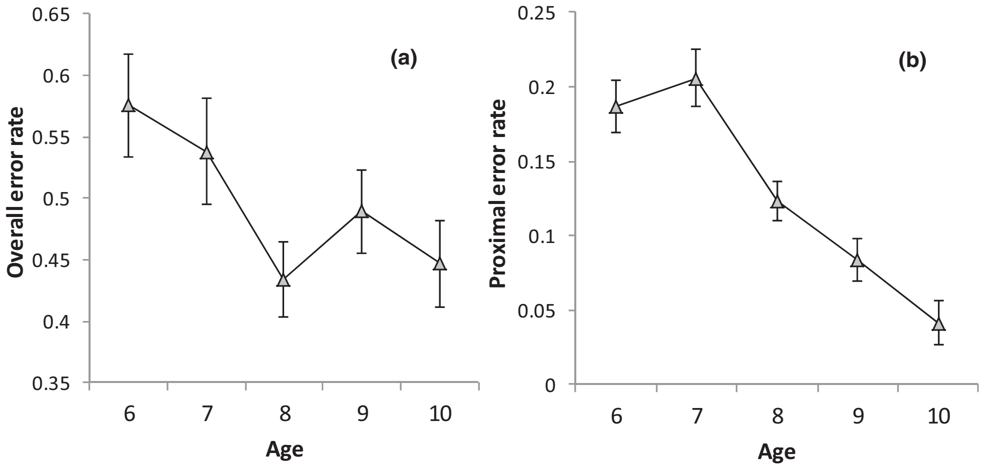

FIGURE 2 Overall error rate (a) and Proximal error rate (b) as a function of Age (Mean ± Standard Error).

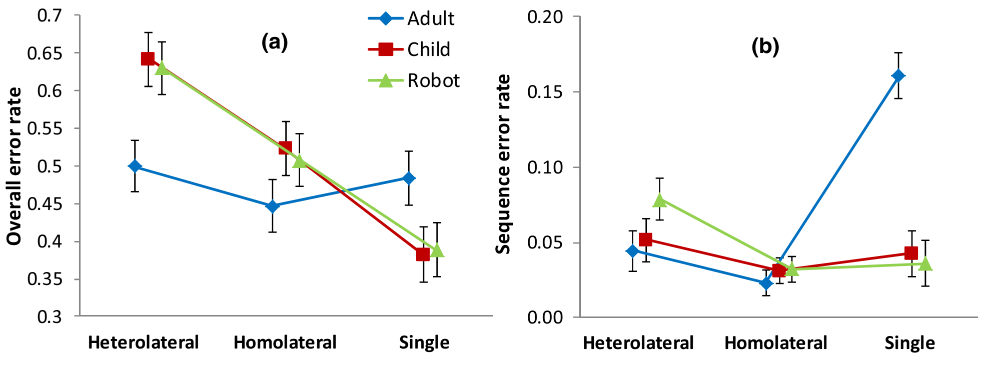

FIGURE 3 Condition × Model interaction (mean ± SE) on Overall error rate (a) and Sequence error rate (b).

were the easiest, Heterolateral actions were the most difficult, and Homolateral had intermediate difficulty. However, this only happened when the model was either a Child ($F(2, 1473) = 22.808$, $p < .001$) or a Robot ($F(2, 1305) = 21.838$, $p < .001$): when the model was an Adult, such effects disappeared ($F(2, 1353) = 1.331$, $p = .264$). Model had a significant effect in the most difficult, Heterolateral condition, where the Adult model showed a large advantage (mean 13.1%, CI = [4.8, 21.3]) over Robot and Child models ($F(2, 1377) = 4.762$, $p = .009$); Model did not reach significance in the Homolateral condition ($F(2, 1377) = 1.316$, $p = .269$) and in the Single-effector condition ($F(2, 1377) = 2.53$, $p = .08$), albeit in the latter the pattern seemed to be exactly the opposite to that found in the Heterolateral condition (children performed *worst* with the Adult model, Figure 3a).

A straightforward interpretation of the *overall* ER is arduous: it is a composite measure reflecting several separate error categories, which in turn show different, even opposite and clashing effects (see Figure 4a). Analyses of *specific* ERs are thus essential.

Error profiles

To provide an overview, Figure 4 shows ERs for specific error categories as a function of Condition (Figure 4a) or Model (Figure 4b). We could not perform statistical comparisons *across* error categories as these are not mutually independent (they are subsets of a constant overall set of trials, so their frequency

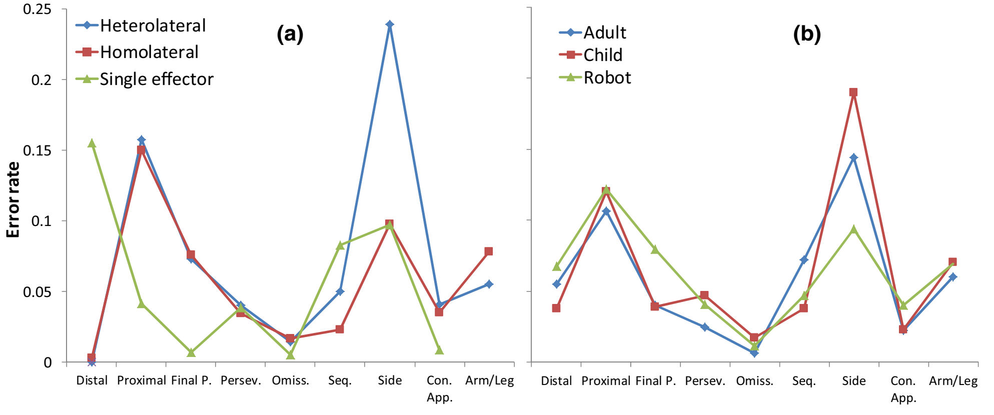

FIGURE 4 Error rate for specific error categories as a function of (a) Condition and (b) Model. Final P., Final Position; Persev., Perseveration; Omiss., Omission; Seq., Sequence; Con. App., Conduite d'Approche. Arm/Leg confusion errors were impossible in Single-effector actions (plot (a)), because these stimuli involved a single limb.

values constrain each other). Only qualitative comparisons are allowed in this case. As to within-error-category comparisons, statistical tests are reported later. Figure 4 shows that, in general, Proximal and Side errors were the most common ones. Interesting within-category differences, according to either Condition (Figure 4a) or Model (Figure 4b) emerged, which are analysed below.

Statistical analyses within specific error categories

Perseveration, Omission, Conduite d'Approche, Arm/Leg (fusion or inversion) ERs failed to show (Bonferroni-corrected) significant effects (Table 2), so we focused on other error categories.

Side errors

The triple interaction had to be removed from GMLM due to convergence failure. The analysis yielded significant effects of Condition ($F(2, 4147) = 58.28, p < .001$) and Age ($F(4, 4147) = 9.864, p < .001$). The rate of Side errors clearly *increased* with Age (Figure 5). According to non-linear regression (Appendix A), the ER began to rise at Age = 7.72 ± 0.89, about 8 years (this paradoxical, error-specific result explains the inversion, at the same Age, of the decreasing trend of the *overall* ER, Figure 2a).

Concerning the effect of Condition (visible in Figure 4a), as expected, Side errors were most frequent when spatial and body-related encoding demands were highest, that is in the Heterolateral condition. In the Homolateral and Single-effector conditions, ERs were much lower, without any differences between them. Thus, when a single side of the body was involved, subjects encoded its laterality relatively easily, no matter whether the action recruited one or two effectors. By contrast, laterality encoding was much less accurate when both body sides were involved (in a crossed pattern).

Albeit marginal after Bonferroni correction ($F(2, 4147) = 4.928, p = .007 > .0055$), the effect of Model on Side errors is worth mentioning, insofar as it involves very large (two-fold) differences in relative frequency (Figure 4b): with the child model, average Side ER was 19% (CI = [14.2, 23.8]), which dropped to 9.4% with the robot model (CI = [4.9, 13.8]); the adult model led to a 14.4% rate (CI = [8.8, 20]). This pattern was confirmed and investigated more deeply in the following analysis of per-limb ERs.

Per-limb rates of Side errors: GEE

The significant disadvantage of Heterolateral with respect to the other two conditions (Figure 5) might be due either to the fact that the processing of the side of one limb is somehow perturbed by the processing

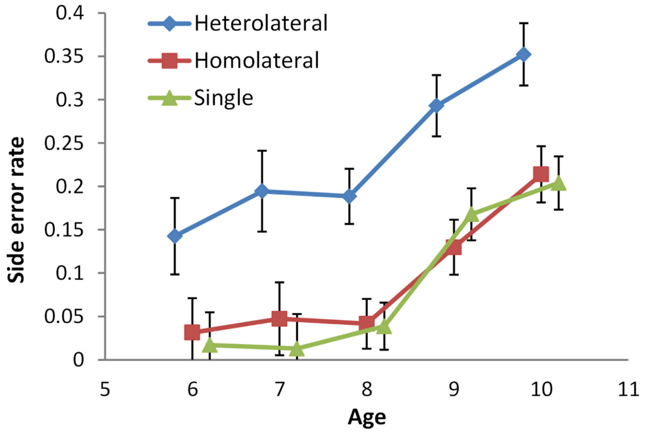

FIGURE 5 Side error rate (Mean ± SE) as a function of Age and Condition.

TABLE 4 Results of GEE on Side error rates.

	Wald Chi-square	*df*	*p*
Age	50.564	4	<.001
Model	7.539	2	.023
Condition	38.299	1	<.001
Age × Model	8.417	8	.394
Age × Condition	25.332	4	<.001
Model × Condition	7.713	2	.021
Age × Model × Condition	19.056	8	.015

of the side of the other ('mutual perturbation' hypothesis, see the 'Measures' Section above), or to the trivial mathematical fact that with two limbs to process, the probability of at least one Side error is higher ('independent processing' hypothesis). The two hypotheses could be disentangled by comparing the per-limb rate of Side errors[1] from the Heterolateral condition with the ER from the other two conditions (Single and Homolateral trials were merged, given that their ERs proved virtually identical, see previous GMLM and Figure 5). The comparison was carried out by a GEE analysis (previously validated in a dedicated Monte-Carlo study) with Age (five classes) and Model (three types) as between-subjects factors, and Condition as within-subjects factor (Adjusted Heterolateral vs. Single/Homolateral Side ERs). Distributions of Side ERs were modelled as Tweedie with log-link.

The results were very interesting. Condition had a significant effect (Table 4): clearly, performance on a limb was strongly affected by the fact that another limb had to be processed on the opposite side, supporting the mutual perturbation hypothesis. The mutual perturbation effect is clearly evident as the gap between the blue (Heterolateral) and the red (Homolateral/Single) plots in Figure 6a. However, the significant Condition × Age interaction showed that such a disadvantage characterized younger children (6–8 years: Wald $\chi^2 = 35.081$, $p < .001$; mutual perturbation effect = 0.064, CI = [0.042, 0.087]) much more so than older children (9–10 years: Wald $\chi^2 = 3.281$, $p = .07$; mutual perturbation effect = 0.028, CI = [−0.01, 0.067]). Indeed, the gap between the blue and the red plot tends to disappear with increasing age (Figure 6a; note that the raw ER, shown in Figure 5, could not detect such a relevant difference and would have wrongly suggested that younger and older children are equally sensitive to the presence of one extra limb to process on the opposite side). Another interesting feature emerged from the

[1]It could not be answered by using raw ERs: see Appendix B.

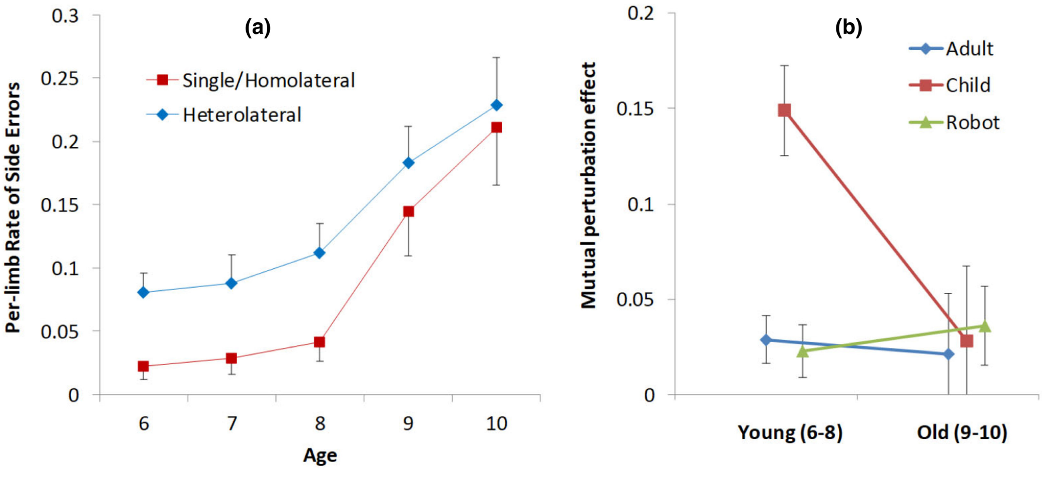

FIGURE 6 (a) Per-limb rate of Side errors (mean ± SE) as a function of Age (years). Red: average between Single and Homolateral error rates (which proved similar in GMLM analyses); blue: Heterolateral error rates. (b) The 'mutual perturbation effect' is the increase in the probability of a Side error on a limb, when another limb on the opposite side is involved in the action. This is shown on the *Y* axis as a function of Age (*X* axis: younger vs. older children) and Model (blue, Adult; red, Child; green, Robot). Table 4 reports GEE analyses' significance levels.

interactions involving Model, especially from the triple interaction (Table 4). This was due to the fact that the Model × Condition interaction was undetectable among older (9–10 years) children (Wald $\chi^2 = 0.665$, $p = .717$), but significant among younger (6–8 years) ones (Wald $\chi^2 = 9.112$, $p = .011$); in turn, the latter interaction arose because the Condition effect was very strong with a Child model (Wald $\chi^2 = 36.962$, $p < .001$), weaker but still significant with an Adult model (Wald $\chi^2 = 9.29$, $p = .002$) and not significant with a Robot model (Wald $\chi^2 = 2.677$, $p = .102$). Figure 6b clarifies the pattern of the triple interaction by showing the Condition effect on the *Y* axis. Hence, the axis directly illustrates the mutual perturbation effect. This was generally very small and close to zero, except for younger children watching a Child model (mean effect = 0.149, CI = [0.101, 0.197]).

To summarize, younger children (6–8 years) made very few per-limb Side errors, but these increased in frequency when another limb had to be moved on the opposite side (and the model was a Child; Figure 6a,b); by contrast, older children (9–10 years) made many per-limb Side errors, but these did *not* increase in frequency when another limb had to be moved on the opposite side (Figure 6a). Thus, the counterintuitive rise of Side errors in older children might be ascribed to factors that affected the two limbs separately.

Distal errors

On the GMLM, Condition had a strong effect ($F(2, 1431) = 76.627$, $p < .001$). As shown in Figure 4a, the percentage of Distal errors was virtually null with Double-effector actions (both Heterolateral and Homolateral) and much higher (15.4%) with Single-effector actions. This apparently paradoxical result is actually trivial: Single-effector actions contained distal components (rotations of the hand/foot) that were not included in Double-effector actions (see Appendix C).

Proximal errors

The full GMLM model did not converge on reliable solutions (yielding very high standard errors of parameter estimates), so we excluded the triple interaction from the model. Results are shown in Figure 4a.

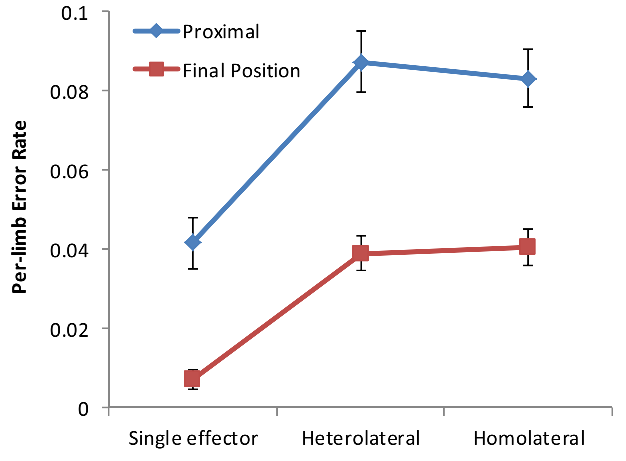

FIGURE 7 Per-limb error rate as a function of Condition; blue, Proximal errors; red, Final Position errors.

Single-effector actions produced *fewer* Proximal errors than double-effector ones (Condition: $F(2, 4147) = 40.543$, $p < .001$). Moreover, a strong effect of Age emerged ($F(4, 4147) = 11.566$, $p < .001$), with the ER being lower in older participants. Non-linear regression (logistic model) estimated the stabilization Age to be 9.75 ± 1.02, which essentially corresponds to the oldest age class tested (those participants had very close-to-zero ERs, Figure 2b).

Like Side errors, Proximal errors also showed a strong disadvantage for Heterolateral with respect to Single-effector actions, raising the question of whether this disadvantage reflected mutual perturbation effects, or independent processing of the two limbs. Hence we computed per-limb ERs, this time for both the Heterolateral and the Homolateral conditions, as they both showed a raw-ER disadvantage with respect to Single-effector actions (Figure 4a). GEE on the per-limb rate of Proximal errors yielded a significant effect of Condition (Wald $\chi^2 = 27.679$, $p < .001$), with both the Heterolateral (Wald $\chi^2 = 28.768$, $p < .001$) and Homolateral (Wald $\chi^2 = 21.914$, $p < .001$) conditions showing a disadvantage with respect to Single (i.e. a mutual perturbation effect; the two estimates were, respectively, 0.046, CI = [0.03, 0.062] and 0.041 CI = [0.025, 0.058]) and without any significant difference between Heterolateral and Homolateral themselves (Wald $\chi^2 = 0.443$, $p = .506$). Thus, having to process another limb, no matter whether on the same or the opposite side (Homolateral or Heterolateral) yielded a reliable cognitive cost (Figure 7, blue plot). This held both for 6/7 year olds (Condition effect: Wald $\chi^2 = 18.878$, $p < .001$) and for 8/9/10 year olds (Wald $\chi^2 = 13.953$, $p < .001$). While Condition × Age did not reach significance (Wald $\chi^2 = 15.361$, $p = .053$), the absolute sizes of the Hetero/Homo versus Single gaps in performance were quite different (0.09 for younger, 0.026 for older children), an expected pattern given that the performance of older children was much closer to the floor.

Final Position errors

On the GMLM analysis, Condition ($F(2, 4131) = 14.321$, $p < .001$) yielded a significant effect due to the usual advantage of Single-effector (0.7% ER) over the other conditions (both at about 7.5% ER), visible in Figure 4a. Such a pattern called for another investigation of the per-limb ER by means of GEE. Like with Proximal errors, the per-limb rate of Final Position errors showed a strong Condition effect (Wald $\chi^2 = 23.531$, $p < .001$), due to an advantage of Single over both Heterolateral (Wald $\chi^2 = 25.879$, $p < .001$; mean = 0.032, CI = [0.023, 0.041]) and Homolateral (Wald $\chi^2 = 24.722$, $p < .001$; mean = 0.033, CI = [0.024, 0.043]) actions and without differences between the latter two (Wald $\chi^2 = 0.427$, $p = .513$). Again, this suggests that Double-effector actions entail mutual perturbation effects (Figure 7, red plot). This held for both 6/7 year olds (Condition effect: Wald $\chi^2 = 7.458$,

p = .024) and for 8/9/10 year olds (Wald $\chi^2 = 19.097$, $p < .001$), with virtually identical absolute sizes (mutual perturbation effects were 0.036 and 0.031, respectively[2]).

Sequence errors

GMLM showed a highly significant Condition × Model interaction ($F(4, 4131) = 6.448$, $p < .001$), which was clearly due to *Adult model*, *Single-effector actions* inducing more Sequence errors than all other combinations (Figure 3b). Indeed, Model had a marked effect in the Single condition ($F(2, 1377) = 16.902$, $p < .001$) while it fell short of significance in the Homolateral ($F(2, 1377) = 0.137$, $p = .872$) and Heterolateral ($F(2, 1377) = 0.98$, $p = .376$) conditions.

GENERAL DISCUSSION

This study investigated the developmental trajectory of imitation in first-grade school children (aged 6–10 years). We manipulated the complexity of meaningless actions (from movements of a single limb to 'Homolateral' movements involving the arm and the leg of the same body side to 'Heterolateral' movements involving arm and leg of opposite sides), and used three models (a human child, a human adult and a child robot, the iCub). We found several interesting patterns of results, which are discussed in separate Sections below.

Importantly, such patterns would have been impossible to detect by means of traditional analyses of accuracy rates: we could observe them thanks to error analysis. This technique also provided us with two specific indices, the per-limb error rate and the 'mutual perturbation effect', which allowed us to understand in what conditions, and to what extent, the processes regarding a limb 'perturbed' the processes concerning the other limb involved in the action. This information was critical, because not all cognitive models are compatible with such mutual perturbation hypothesis. For instance, the use of limited-capacity storage systems (like short-term memory, STM) predicts mutual perturbation between the stored elements, while the involvement of systems that process information in parallel (as, e.g. visual analysis) does not. Therefore, the above indices offered useful theoretical cues towards plausible cognitive explanations of the present findings. These are discussed below.

Role of model

We will start with the least clear set of results. Relying on the previous literature (see 'Predictions' Section in the Introduction), we expected that an adult model should yield a better imitation performance than a peer or a robot, given that we used new, meaningless actions as stimuli. Indeed, this material typically induces reliance on adult models by children (Taylor et al., 1991). When analysing the overall error rate (overall ER), this prediction was only partially confirmed, as such a pattern characterized only the most difficult, Heterolateral actions, and to a lesser extent, Homolateral ones. The simplest actions, involving a single limb, showed, if anything, a paradoxical *disadvantage* if they were imitated from an adult model (Figure 3a), mostly due to Sequence errors (Figure 3b). We have no straightforward explanation for the latter result. While most error categories failed to show obvious differences across models, we found another puzzling pattern in the context of Side errors (the most frequent ones). In these, a specific advantage was found for both the adult and the robot model over the child model, in younger (6–8 years old) children and for Heterolateral actions (Figure 6b). One possibility is that both the adult and the robot model increased young children's attention, albeit for different reasons: the adult model might have done so as an effect of perceived authoritativeness and/or through motivational factors (Ottoboni et al., 2013;

[2]We could not test the Condition × Age term directly, as this yielded a singular Hessian matrix.

Over & Carpenter, 2012; Taylor et al., 1991) and the robot model by involving children in a game-like interaction (as shown in a wide range of studies on autistic children, e.g. Michaud et al., 2003, Michaud & Théberge-Turmel, 2002, or in healthy participants in a physiotherapy session; Brooks & Howard, 2012). However, we do not have an explanation as to why this pattern was found only for Side errors in Heterolateral actions.

Role of age

Common sense, as well as two of the main theories of action imitation (Bekkering et al., 2000; Tessari & Rumiati, 2004), predicted an improvement of performance with age, assuming that the general maturation of the cognitive system, and particularly, the expansion of working memory (WM) capacity, would drive such an effect. Overall performance improved with age but showed an uncertain pattern, with a sudden inversion of this trend between ages 8 and 9 and again an improvement between ages 9 and 10 (Figure 2a). Some error categories showed a coherent improvement (e.g. Proximal, Figure 2b). However, exactly the *opposite* held true for the most frequent error category, Side errors. These were stable in frequency (and virtually absent in two sub-categories) at ages 6 to 8; their frequency then suddenly rose at age 9, and increased even further in 10 year olds (Figure 5). This unexpected increase accounts for the presence of the strange 'bump' in the overall ER between ages 8 and 9 (Figure 2a), but is by itself very puzzling: none of the theories of imitation we are aware of can explain why Side ER increased with age. We believe the most plausible explanation for this finding is that older children might experience some form of interference by an 'anatomical' imitation strategy, which would be absent from the cognitive repertoire of younger children (see later), thus leading to the paradoxical advantage of the latter group. By 'anatomical strategy' we mean that older children might sometimes have assumed the model's point of view and, for example, if the model moved his right hand, they also moved their right hand. This strategy clashes with the instructions, which explicitly required participants to imitate 'as if they were in front of a mirror': thus, they should have moved their *left* hand.

More on Side errors

The above explanation should be extended to account for other evidence. Side errors were most frequent in the Heterolateral condition, which is hardly surprising given that Heterolateral actions are supposed to be the most difficult items. However, such a disadvantage was not due to the trivial fact that two limbs had to be moved rather than one, but rather, to the fact that those two limbs were in a crossed pattern—left arm plus right leg or vice versa. Indeed, Homolateral actions also involved two limbs, but these were on the same side of the body, and Side ER for them was as low as in the baseline condition, the one that involved a single limb. We believe such a pattern of results can be interpreted by reasoning on the number of spatial representations (codes) that children internally produce to specify the laterality of the action. Clearly, the more codes are produced, the more likely it is that at least one of them is wrong (i.e. the higher the probability of a Side error to occur). We started from the hypothesis that children produced separate laterality codes for arm and leg. If this had been the case, Double-effector movements would have shown more Side errors than Single-effector movements. Thus, for instance, a Homolateral trial involving the left arm and the left leg would generate two separate codes, 'arm = left' and 'leg = left', and errors would be more frequent than in the baseline condition involving a single code, (e.g.) 'arm = left'. Results contradicted this prediction: the Homolateral and the Single-effector conditions yielded virtually identical Side ERs (Figures 4a and 5)—it was only the Heterolateral condition that yielded a higher Side ER.

We believe the most plausible explanation is that in both the Single-effector and Homolateral-effectors conditions, children produced *only one* laterality code, attached either to the single limb involved (Single condition) or to the arm-leg pair (Homolateral condition), e.g. 'arm+leg = left', thus having the same probability of making a Side error in both cases. By contrast, in the Heterolateral condition they could

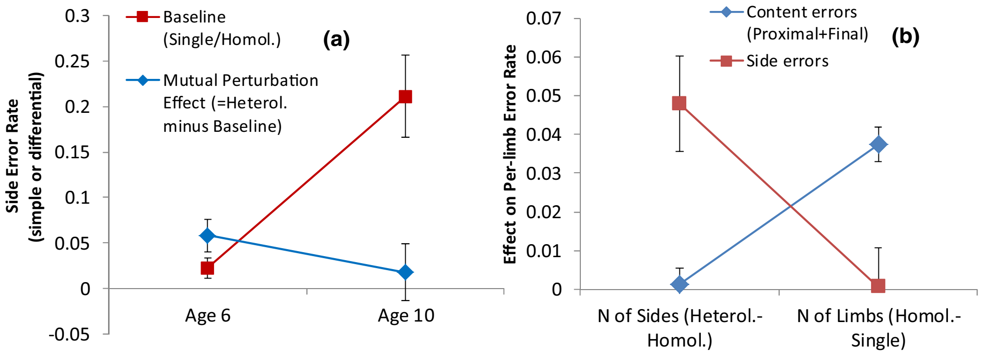

FIGURE 8 The two double dissociations found in this study. (a) Red plot: the raw rate of Side errors in the 'Baseline' conditions (Single and Homolateral, averaged) *increased* with Age, an effect attributed to 'anatomical' imitation in older children; blue plot: the 'mutual perturbation effect' (i.e. the surplus of per-limb Side errors found in the Heterolateral with respect to the Baseline conditions) *decreased* with Age—an effect attributed to higher WM capacity in older children. Only extreme ages from our sample (6, 10) are shown: see Figure 6a for other ages. (b) The effects of two stimulus variables (the number of body sides, and the number of limbs, involved in the action) on two specific per-limb error rates ('Side' errors in red and 'Content' errors in blue) are shown. *N* of sides changed from 1 to 2 between the Homo- and the Heterolateral conditions (with *N* of limbs being constant, 2, arm and leg); *N* of limbs changed from 1 to 2 between the Single and the Homolateral conditions (with *N* of sides being constant, 1). The *Y* axis is the mean (±SE) change produced in error rate (zero = no change; positive = increase) by the two stimulus variables. Doubling the *N* of sides produced an increase in Side error rate, but no change in Content error rate, and doubling the *N* of limbs produced exactly the opposite pattern. This cross-over schema suggests the use of separate STM systems for encoding the sides and the motor contents of the actions. All four differences of the cross-over pattern were highly significant ($df = 173$, $p < .001$; Content vs. Side errors: $t = -8.08$ for the effect of *N* of sides, $t = 7.38$ for the effect of *N* of limbs; contrast between the two effects, *N* sides vs. *N* limbs: $t = -6.99$ for Content errors, $t = 6.61$ for Side errors).

not avoid generating two separate codes, one for the side of the arm and another for the side of the leg (e.g. 'arm = left' and 'leg = right'); by using two separate codes, error probability unavoidably increased.

Another intriguing piece of evidence was the following. When processing the side of a limb, the presence of another limb to be moved on the *opposite* side of the body increased the probability of a Side error—what we called a 'mutual perturbation effect'. Critically, this effect characterized the performance of younger children (6–8 years) and tended to disappear in older ones (9–10 years)—see Figure 6a. Why did the mutual perturbation disappear with increasing age? We have two possible explanations. On the first, the mutual perturbation occurs because two laterality codes (e.g. 'left' for the arm and 'right' for the leg) compete for the same limited cognitive resources, like the capacity of a STM storage. With increasing age, STM capacity would also increase, and with more storage space, less competition would occur. In the second explanation, the mutual perturbation occurs because the two laterality codes *specifically* interfere with each other, given that they specify opposite sides, and younger children might be less able to manage such conflicting information. In any case, one of the components of WM (Baddeley, 2010) would be involved, either storage (STM) or executive functions (once collectively called 'Central Executive', but see Logie, 2016), which are held responsible for managing conflicting representations (e.g. Ambrosi et al., 2019). Limited WM in children has indeed been documented (e.g. Burnett Heyes et al., 2012) and assumed to be crucial in imitation (Wohlschläger et al., 2003; specific work confirmed the limits of the STM component in this population: Gathercole, 1999; Isaacs & Vargha-Khadem, 1989; Wilson et al., 1987).

To summarize, Side errors showed a 'double dissociation' between younger and older children (Shallice, 1988; Figure 8a). Younger children experienced some mutual perturbation (quantifiable as the gap between the red and the blue plot, Figure 6a) between the laterality codes of limbs belonging to opposite body sides, possibly due to their WM limits; however, they were *not* hindered by an anatomical imitation strategy, which is likely to be still absent from their cognitive repertoire. Older children showed precisely the opposite pattern: they did *not* show any mutual perturbation when encoding two sides (red and blue plots converge), as maturation likely provided them with sufficient WM abilities; however, they sometimes applied an anatomical imitation strategy, which worsened their side-encoding performance (both curves, red and blue, increase with age).[3] Note that the difficulty experienced by younger children in the Heterolateral condition almost only affected trials in which the model was a child (Figure 6b). We already discussed the possibility that the adult and the robot models elicit more attention/interest in this population. In this view, a greater attentional involvement might have masked the limits in STM capacity and/or conflict-managing abilities, which would have emerged with a less engaging child model.

While intriguing and based on reliable statistical results, the above reconstruction is certainly *post-hoc* and deserves further investigation. As a first step in that direction, we checked whether the predictions of our proposed accounts actually correspond to the empirical data, by means of extensive Monte-Carlo simulations. The results of this study are still preliminary, so we reported them in Appendix B; however, given their insight, they are worth briefly mentioning here. This study confirmed that the pattern of Side errors in younger children is compatible with the predictions of a marked STM limit (blue curve in Figure B1) or a significant incapacity to manage the conflict between side labels (red curve). It also confirmed that older children's performance instead is compatible with the predictions of an anatomical imitation strategy and not at all with STM/conflict-managing limits. However, interestingly, the performance of older children showed the maximum compatibility *not* with the hypothesis that they were applying an anatomical imitation strategy to the whole body (green line), but rather with the hypothesis that such a strategy was applied independently to the arm and the leg (violet line). This speculative possibility needs further theoretical and empirical work.

'Content' (Proximal and Final Position) errors

We will discuss Proximal and Final Position errors together, as they both led to a similar pattern of results (Figure 7), which in turn was different from that found with Side errors. Proximal and Final Position errors concern the motor features of the action—what we will henceforth refer to as its 'content' (for instance, the content of the action depicted in Figure 1c is that (i) the arm must be *lifted forwards* and (ii) stopped at *shoulder level*, while the leg must be (iii) *lifted laterally* and (iv) stopped at *45°*). So we will refer to the union between those categories as 'Content' errors. Like Side errors, also Content errors were much more frequent in the Heterolateral than in the baseline, Single-limb condition (Figure 4a). However, as expected, this time, just the *number* of limbs counted and not their side. Indeed, the Homolateral and the Heterolateral conditions yielded virtually identical ERs. While the *side* of, for example, the left arm and the left leg could well be 'merged' into a single laterality code (e.g. 'left limbs'), thus making the Side ER drop for the Homolateral condition, this was, of course, impossible when the content of the movement had to be encoded, because the content (including all of its parameters, angles, targets, etc.) was always different between arm and leg.

With Content errors, both younger and older children showed significant perturbation of the processing concerning one limb by the presence of another limb in the action. Indeed, in both conditions with two effectors (Heterolateral and Homolateral), younger as well as older children had a higher per-limb ER than that recorded in the Single-effector condition (Figure 7). Again, STM limitation and/or conflict-managing difficulties may be plausible explanations for such a pattern.

Separate STM systems for action content and side?

Much work has demonstrated a specific STM system dedicated to storing actions (Ottoboni et al., 2021; Rumiati & Tessari, 2002; Tessari & Rumiati, 2002). Another statistically reliable (albeit unexpected) finding

[3]A proof that older children's Side errors were not caused by WM limits, but rather by a wrong imitation strategy, is that those errors were equally frequent in the Single-effector condition, where WM demands are minimal, and in the Heterolateral one, where WM demands are maximal (by 'equally frequent' we refer to the per-limb ER, Figure 6a: this index partials out the between-conditions difference in the number of limbs moved).

of the present work suggests the existence of an even subtler distinction. The idea is that different STM systems might be involved in storing the side(s) and the content of the action. By playing the devil's advocate, suppose that a single STM system were used to store both information on the side(s) and the content of the action. In this scenario, the only critical variable would be the overall load, so whether one increases the load by increasing the number of sides, or by increasing the content of the action, the effect would be the same: the frequency of errors would rise—errors of any nature, so both Side ER and Content (Proximal/Final Position) ER would increase. Effects would all be non-specific. By contrast, the effects would be specific if separate STM for side(s) and content were used: increasing the number of sides would cause an increase in Side errors, but not in Content errors; increasing the content of the action would cause an increase in Content errors, but not in Side errors. In other words, the predicted pattern would be a cross-over or double dissociation. Luckily, our experimental design allowed for a test of such a prediction, because it varied the number of sides and the content (number of limbs) independently: Homolateral and Heterolateral actions differed for the number of sides involved (1 vs. 2), but not for the content (both conditions included simultaneous movements of one arm and one leg, and the movements were perfectly matched across conditions); Homolateral and Single-effector actions differed for the content (1 vs. 2 moved limbs) but not for the number of sides involved (1 in both cases: Homolateral actions involved both limbs of the same side). Results were consistent with the cross-over prediction (Figure 8b): increasing the number of sides *exclusively* increased the rate of Side errors (left half of Figure 8b), and increasing the number of limbs *exclusively* increased the rate of Content (Proximal/Final Position) errors (right half of Figure 8b). Therefore, the hypothesis of separate STM systems for the side(s) and the motor contents of the action was confirmed.

CONCLUSION

The most intriguing hypotheses of the present study arose from unexpected but reliable results provided by the analysis of error types. Traditional analyses of simple accuracy scores would have been completely blind to those features. These hypotheses concern (i) the possibility that separate STM subsystems are used in action imitation, one for encoding the side(s) of the limb(s) involved in the action and another for encoding the motor features (the 'content') of the action, and (ii) the putative roles of 'mirror' versus 'anatomical' imitation strategies as a function of children's age.

As to the putatively separate STM systems, given that the present experiment had not been explicitly planned to tackle this issue, further research is necessary to clarify their exact nature, and particularly, whether or not they correspond to Baddeley (2010) visuospatial sketchpad, for storing the limbs' sides, and to the 'motor STM' (Ottoboni et al., 2021; Rumiati & Tessari, 2002; Tessari & Rumiati, 2002), for storing the other motor content of the action. These speculative proposals will be the subject of future work.

As to imitation strategies, these were brought into play in order to explain the surprising increase in the frequency of Side errors with age. We hypothesized a shift from purely 'mirror' (or 'specular') imitation in younger (6–8 years old) children, to the presence of some 'anatomical' imitation in older (9–10 years old) children. Specular imitation—a type of allocentric imitation (Bianchi et al., 2014)—is more natural than anatomical (egocentric) imitation until age 10 (Bergès & Lézine, 1963; Brass et al., 2000; Gleissner et al., 2000; Schofield, 1976; Wapner & Cirillo, 1968). Even adults show the tendency to imitate specularly in many circumstances because anatomical imitation is a more demanding mental operation: it requires inhibiting the automatic tendency to mirror the model's movements, as well as an additional spatial transformation of the perceived movements, from the model's body to the imitator's body (see data from patients with frontal lesions: Chiavarino et al., 2007, or from healthy adults: Avikainen et al., 2003; Mengotti et al., 2012, 2013). Hence, anatomical imitation is likely to require both WM resources and accurate knowledge of one's own body.

Regarding WM, previous work supports the view that WM plays a crucial role in imitation (Cubelli et al., 2000; Ottoboni et al., 2021; Rumiati & Tessari, 2002; Tessari & Rumiati, 2002; Wohlschläger et al., 2003) and has a capacity, which strongly depends on age. Several authors reported a sizable increase of WM resources between ages 4 and 14 (Gathercole, 1999; Isaacs & Vargha-Khadem, 1989;

Wilson et al., 1987), with a steep gradient between 4 and 8, and a shallower gradient between 8 and 12 (Nelson, 1995). While we did not explicitly measure WM in the present study, indirect evidence that WM resources increased through age is nonetheless available: we refer to the improvement of *overall* imitation performance through age classes (Figure 2), and especially to the fact that with increasing age, performance became less and less sensitive to the number of sides to be processed (Figure 6a). Some preliminary work in which WM functioning was explicitly simulated (in two components, STM and conflict-managing ability, see Appendix B, Figure B1) confirmed that the change in performance between younger and older children is compatible with an increase in WM capacity.

Regarding body knowledge, the development of the multisensory processing that underlies body perception seems to follow a long-time course (Begum Ali et al., 2014; Bremner et al., 2013; Cowie et al., 2013; Nardini et al., 2013; Pagel et al., 2009). As demonstrated with the rubber hand illusion paradigm, 4- to 9-year-old children rely more on visual than on proprioceptive information when localizing their own hands; the balance between visual and proprioceptive inputs, which is typical of adults, is reached at age 10–11 (Cowie et al., 2016). Another study showed how the representation of one own's body develops progressively until age 8, and beliefs concerning the body of others mature later, from age 8 to 10 (Assaiante et al., 2014). Moreover, 8 is when children develop proper semantic knowledge of both the upper and lower body parts (Auclair & Jambaqué, 2015). Future work might be crucial to investigate the role of different body representations: body schema (see Corradi Dell'Acqua & Rumiati, 2007, for a review), body image (Sirigu et al., 1991) and structural description of the body (Ottoboni et al., 2005; Schwoebel & Coslett, 2005; Tessari et al., 2010) that are known to dissociate (Tessari & Ottoboni, 2022) and have different developmental trajectories.

In summary, 9/10-year-old children seem to possess both of the putative prerequisites for performing anatomical imitation: they have the required WM resources and sufficient knowledge of the human body. This would explain why, in our experiment, only 9/10 year olds showed the increase in Side errors which we attributed to anatomical imitation. Anyway, even taking for granted that 9/10 year olds can imitate anatomically, why did they do so in our experiment, given that anatomical imitation is a more difficult cognitive operation (and was not required by the instructions)? Are 9/10 year olds beginning to 'experience' and 'test' their new body knowledge? We leave this as an open and critical question deserving of further research.

AUTHOR CONTRIBUTIONS

Giovanni Ottoboni: Conceptualization; data curation; investigation; methodology; project administration; writing – original draft. **Alessio Toraldo:** Conceptualization; data curation; development and validation of measures; statistical analysis; theoretical analysis of the results; writing – original draft. **Riccardo Proietti:** Investigation; writing – original draft. **Angelo Cangelosi:** Conceptualization; resources; writing – original draft. **Alessia Tessari:** Conceptualization; data curation; investigation; methodology; project administration; resources; software; supervision; writing – original draft.

ACKNOWLEDGEMENTS

The authors thank Justin Harris (University of Sydney) for revising the English. The authors also thank Eleonora Calimazzo and Elisabetta Dondi for supporting the participating children during the experiment, Prof. Andrea Ceciliani (University of Bologna) for having facilitated the interaction with the schools, and Gabriele Russo (University of Bologna) for his help in recording part of the stimuli. Finally, the authors thank all the families of the children who agreed to participate in the study.

Open Access Funding provided by Universita degli Studi di Bologna within the CRUI-CARE Agreement.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the finding for this study are available from the corresponding authors, and can be downloaded from https://tinyurl.com/alessiotoraldo, or from https://osf.io/daewb/?view_only=bb85db7595a54082bf6900d1aef5064e.

ETHICAL APPROVAL

Ethical approval was granted by the Bioethics Committee of the University of Bologna on the 7 May 2013, following the Declaration of Helsinki, and parents/tutors provided written informed consent for the children to participate in the study.

ORCID

Giovanni Ottoboni https://orcid.org/0000-0002-7649-6642
Alessio Toraldo https://orcid.org/0000-0001-6480-4751
Riccardo Proietti https://orcid.org/0000-0002-2490-3103
Angelo Cangelosi https://orcid.org/0000-0002-4709-2243
Alessia Tessari https://orcid.org/0000-0002-6616-2454

REFERENCES

Ambrosi, S., Servant, M., Blaye, A., & Burle, B. (2019). Conflict processing in kindergarten children: New evidence from distribution analyses reveals the dynamics of incorrect response activation and suppression. *Journal of Experimental Child Psychology*, *177*, 36–52. https://doi.org/10.1016/j.jecp.2018.06.006

Assaiante, C., Barlaam, F., Cignetti, F., & Vaugoyeau, M. (2014). Body schema building during childhood and adolescence: A neurosensory approach. *Neurophysiologie Clinique/Clinical Neurophysiology*, *44*(1), 3–12.

Auclair, L., & Jambaqué, I. (2015). Lexical-semantic body knowledge in 5- to 11-year-old children: How spatial body representation influences body semantics. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *21*(4), 451–464. https://doi.org/10.1080/09297049.2014.912623

Avikainen, S., Liuhanen, S., Schürmann, M., & Hari, R. (2003). Enhanced extrastriate activation during observation of distorted finger postures. *Journal of Cognitive Neuroscience*, *15*(5), 658–663.

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), R136–R140. https://doi.org/10.1016/j.cub.2009.12.014

Bauer, P. J. (1992). Holding it all together: How enabling relations facilitate young children's event recall. *Cognitive Development*, *7*(1), 1–28. https://doi.org/10.1016/0885-2014(92)90002-9

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Begum Ali, J., Cowie, D., & Bremner, A. J. (2014). Effects of posture on tactile localization by 4 years of age are modulated by sight of the hands: Evidence for an early acquired external spatial frame of reference for touch. *Developmental Science*, *17*(6), 935–943.

Bekkering, H., Wohlschlager, A., & Gattis, M. (2000). Imitation of gestures in children is goal-directed. *The Quarterly Journal of Experimental Psychology: Section A*, *53*(1), 153–164.

Benton, A. L. (1959). *Right-left discrimination and finger localization: Development and pathology*. Paul B. Hoeber.

Bergès, J., & Lézine, I. (1963). *Test d'imitation de gestes: Techniques d'exploration du appor corporel et des praxies chez l'enfant de 3 à 6 ans. [Test of imitation of gestures: Methods for plotting body image and movements of children from 3 to 6 years of age.]*. (pp. X, 128). Masson & Cie.

Bianchi, I., Savardi, U., Burro, R., & Martelli, M. F. (2014). Doing the opposite to what another person is doing. *Acta Psychologica*, *151*, 117–133. https://doi.org/10.1016/j.actpsy.2014.06.003

Brass, M., Bekkering, H., Wohlschläger, A., & Prinz, W. (2000). Compatibility between observed and executed finger movements: Comparing symbolic, spatial, and imitative cues. *Brain and Cognition*, *44*(2), 124–143.

Bremner, A. J., Hill, E. L., Pratt, M., Rigato, S., & Spence, C. (2013). Bodily illusions in young children: Developmental change in visual and proprioceptive contributions to perceived hand position. *PLoS One*, *8*(1), e51887.

Brooks, D. A., & Howard, A. M. (2012). Quantifying upper-arm rehabilitation metrics for children through interaction with a humanoid robot. *Applied Bionics and Biomechanics*, *9*(2), 157–172. https://doi.org/10.1155/2012/978498

Burnett Heyes, S., Zokaei, N., van der Staaij, I., Bays, P. M., & Husain, M. (2012). Development of visual working memory precision in childhood. *Developmental Science*, *15*(4), 528–539.

Buxbaum, L. J., & Randerath, J. (2018). Limb apraxia and the left parietal lobe. *Handbook of Clinical Neurology*, *151*, 349–363. https://doi.org/10.1016/B978-0-444-63622-5.00017-6

Carmo, J. C., & Rumiati, R. I. (2009). Imitation of transitive and intransitive actions in healthy individuals. *Brain and Cognition*, *69*(3), 460–464.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*, *21*(2), 315–330. https://doi.org/10.1016/S0163-6383(98)90009-1

Carpenter, M., Uebel, J., & Tomasello, M. (2013). Being mimicked increases prosocial behavior in 18-month-old infants. *Child Development*, *84*(5), 1511–1518.

Chiavarino, C., Apperly, I. A., & Humphreys, G. W. (2007). Exploring the functional and anatomical bases of mirror-image and anatomical imitation: The role of the frontal lobes. *Neuropsychologia*, *45*(4), 784–795. https://doi.org/10.1016/j.neuropsychologia.2006.08.007

Chow, V., Poulin-Dubois, D., & Lewis, J. (2008). To see or not to see: Infants prefer to follow the gaze of a reliable looker. *Developmental Science*, *11*(5), 761–770.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Conti, D., Di Nuovo, S., Buono, S., Trubia, G., & Di Nuovo, A. (2015). Use of robotics to stimulate imitation in children with autism Spectrum disorder: A pilot study in a clinical setting. In *2015 24th IEEE international symposium on robot and human interactive communication (RO-MAN)* (pp. 1–6). IEEE.

Corradi Dell'Acqua, C., & Rumiati, R. I. (2007). What the brain knows about the body: Evidence for dissociable representations. In F. Santoianni & C. Sabatano (Eds.), *Brain development in learning environments: Embodied and perceptual advancements* (pp. 50–64). Cambridge Scholar Publishing.

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.

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

Cubelli, R., Marchetti, C., Boscolo, G., & Della Sala, S. (2000). Cognition in action: Testing a model of limb apraxia. *Brain and Cognition*, *44*(2), 144–165.

Cubelli, R., Bartolo, A., Nichelli, P., & Della Sala, S. (2006). List effect in apraxia assessment. *Neuroscience Letters*, *407*(2), 118–120.

Della Sala, S., Faglioni, P., Motto, C., & Spinnler, H. (2006). Hemisphere asymmetry for imitation of hand and finger movements. *Goldenberg's Hypothesis Reworked. Neuropsychologia*, *44*(8), 1496–1500.

Dickerson, K., Gerhardstein, P., Zack, E., & Barr, R. (2012). Age-related changes in learning across early childhood: A new imitation task. *Developmental Psychobiology*, *55*, 719–732. https://doi.org/10.1002/dev.21068

Duquette, A., Michaud, F., & Mercier, H. (2008). Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. *Autonomous Robots*, *24*(2), 147–157.

Fiske, S. T. (2010). *Social beings: Core motives in social psychology* (2nd ed.). Wiley.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, *3*(11), 410–419.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177–190.

Gleissner, B., Bekkering, H., & Meltzoff, A. N. (2000). Childrens coding of human action: Cognitive factors influencing imitation in 3-year-old. *Developmental Science*, *3*(4), 405–414.

Gordon, H. (1923). Hand and ear tests. *British Journal of Psychology*, *13*, 283–300.

Haun, D. B., & Tomasello, M. (2011). Conformity to peer pressure in preschool children. *Child Development*, *82*(6), 1759–1767.

Heyes, C. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences*, *5*(6), 253–261.

Heyes, S. B., Zokaei, N., & Husain, M. (2016). Longitudinal development of visual working memory precision in childhood and early adolescence. *Cognitive Development*, *39*, 36–44.

Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Animal Cognition*, *8*, 164–181.

Hoehl, S., Keupp, S., Schleihauf, H., McGuigan, N., Buttelmann, D., & Whiten, A. (2019). 'Over-imitation': A review and appraisal of a decade of research. *Developmental Review*, *51*, 90–108.

Howard, C. M., Smith, L. L., Coslett, H. B., & Buxbaum, L. J. (2019). The role of conflict, feedback, and action comprehension in monitoring of action errors: Evidence for internal and external routes. *Cortex*, *115*, 184–200.

Howard, E. M., Ropar, D., Newport, R., & Tunçgenç, B. (2021). Social context facilitates visuomotor synchrony and bonding in children and adults. *Scientific Reports*, *11*(1), 1–14.

Huang, C.-T., & Charman, T. (2005). Gradations of emulation learning in infants' imitation of actions on objects. *Journal of Experimental Child Psychology*, *92*(3), 276–302. https://doi.org/10.1016/j.jecp.2005.06.003

Hurley, S., & Chater, N. (2005). *Perspectives on imitation: From neuroscience to social science: Vol. 2: Imitation, human development, and culture* (pp. xi, 547). MIT Press.

Isbell, E., Fukuda, K., Neville, H. J., & Vogel, E. K. (2015). Visual working memory continues to develop through adolescence. *Frontiers in Psychology*, *6*, 696.

Isaacs, E. B., & Vargha-Khadem, F. (1989). Differential course of development of spatial and verbal memory span: A normative study. *British Journal of Developmental Psychology*, *7*(4), 377–380.

Jaswal, V. K., & Malone, L. S. (2007). Turning believers into skeptics: 3-year-olds' sensitivity to cues to speaker credibility. *Journal of Cognition and Development*, *8*(3), 263–283.

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science*, *17*(9), 757–758. https://doi.org/10.1111/j.1467-9280.2006.01778.x

Jones, S. S. (2007). Imitation in infancy: The development of mimicry. *Psychological Science*, *18*(7), 593–599. https://doi.org/10.1111/j.1467-9280.2007.01945.x

Jones, S. S. (2009). The development of imitation in infancy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2325–2335. https://doi.org/10.1098/rstb.2009.0045

Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior*, *27*(3), 145–162.

Laland, K. N. (2004). Social learning strategies. *Animal Learning & Behavior*, *32*(1), 4–14.

Logie, R. H. (2016). Retiring the central executive. *Quarterly Journal of Experimental Psychology*, *69*, 10. https://doi.org/10.1080/17470218.2015.1136

Longo, M. R., Kosobud, A., & Bertenthal, B. I. (2008). Automatic imitation of biomechanically possible and impossible actions: Effects of priming movements versus goals. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(2), 489–501. https://doi.org/10.1037/0096-1523.34.2.489

Lyons, D. E., Damrosch, D. H., Lin, J. K., Macris, D. M., & Keil, F. C. (2011). The scope and limits of overimitation in the transmission of artefact culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1158–1167.

McCloskey, M. (2003). Beyond task dissociation logic: A richer conception of cognitive neuropsychology. *Cortex*, *39*(1), 196–202.

McGuigan, N., Makinson, J., & Whiten, A. (2011). From over-imitation to super-copying: Adults imitate causally irrelevant aspects of tool use with higher fidelity than young children. *British Journal of Psychology*, *102*(1), 1–18.

McGuigan, N., & Whiten, A. (2009). Emulation and "overemulation" in the social learning of causally opaque versus causally transparent tool use by 23- and 30-month-olds. *Journal of Experimental Child Psychology*, *104*(4), 367–381. https://doi.org/10.1016/j.jecp.2009.07.001

McGuigan, N., Whiten, A., Flynn, E., & Horner, V. (2007). Imitation of causally opaque versus causally transparent tool use by 3-and 5-year-old children. *Cognitive Development*, *22*(3), 353–364.

Meltzoff, A. N. (1988). Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental Psychology*, *24*(4), 470–476. https://doi.org/10.1037/0012-1649.24.4.470

Meltzoff, A. N. (1996). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Annual Progress in Child Psychiatry and Child Development*, *31*(5), 838–850. https://doi.org/10.1037/0012-1649.31.5.838

Meltzoff, A. N., & Moore, M. K. (1997). Explaining facial imitation: A theoretical model. *Infant and child development*, *6*(3-4), 179–192.

Meltzoff, A. N., Kuhl, P. K., Movellan, J., & Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science*, *325*(5938), 284–288. https://doi.org/10.1126/science.1175626

Mengotti, P., Corradi-Dell'Acqua, C., Negri, G. A., Ukmar, M., Pesavento, V., & Rumiati, R. I. (2013). Selective imitation impairments differentially interact with language processing. *Brain*, *136*(8), 2602–2618.

Mengotti, P., Corradi-Dell'Acqua, C., & Rumiati, R. I. (2012). Imitation components in the human brain: An fMRI study. *NeuroImage*, *59*(2), 1622–1630.

Michaud, F., Duquette, A., & Nadeau, I. (2003). *Characteristics of mobile robotic toys for children with pervasive developmental disorders*. SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483) (Vol. 3, pp. 2938-2943).

Michaud, F., & Théberge-Turmel, C. (2002). Mobile robotic toys and autism. In *Socially intelligent agents* (pp. 125–132). Springer.

Nagy, E., Compagne, H., Orvos, H., Pal, A., Molnar, P., Janszky, I., Loveland, K. A., & Bardos, G. (2005). Index finger movement imitation by human neonates: Motivation, learning, and left-hand preference. *Pediatric Research*, *58*(4), 749–753.

Nardini, M., Begus, K., & Mareschal, D. (2013). Multisensory uncertainty reduction for hand localization in children and adults. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 773–787.

Nelson, C. A. (1995). The ontogeny of human memory: A cognitive neuroscience perspective. *Developmental Psychology*, *31*(5), 723–738. https://doi.org/10.1037/0012-1649.31.5.723

Newcombe, F., & Marshall, J. C. (1973). Stages in recovery from dyslexia following a left cerebral abscess. *Cortex*, *9*(3), 329–332.

Nielsen, M. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology*, *42*(3), 555–565. https://doi.org/10.1037/0012-1649.42.3.555

O'Sullivan, E. P., Bijvoet-van den Berg, S., & Caldwell, C. A. (2018). Automatic imitation effects are influenced by experience of synchronous action in children. *Journal of Experimental Child Psychology*, *171*, 113–130.

Ottoboni, G., Calimazzo, E., Ceciliani, A., & Tessari, A. (2013). Il rapporto funzionale tra l'apprendimento per imitazione e la metrica corporea: Uno studio pilota con bambini. *Giornale Italiano di Psicologia*, *4*(2013), 839–850. https://doi.org/10.1421/76949

Ottoboni, G., Ceciliani, A., & Tessari, A. (2021). The effect of structured exercise on short-term memory subsystems: New insight on training activities. *International Journal of Environmental Research and Public Health*, *18*(14), 7545. https://doi.org/10.3390/ijerph18147545

Ottoboni, G., Sciulli, I., & Tessari, A. (2018). L'apprendimento per imitazione: Modelli generali e specificità. *Infanzia*, *1*, 37–42.

Ottoboni, G., Tessari, A., Cubelli, R., & Umiltà, C. (2005). Is handedness recognition automatic? A study using a simon-like paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(4), 778–789. https://doi.org/10.1037/0096-1523.31.4.778

Over, H., & Carpenter, M. (2012). Putting the social into social learning: Explaining both selectivity and fidelity in children's copying behavior. *Journal of Comparative Psychology*, *126*(2), 182.

Over, H. (2016). The origins of belonging: Social motivation in infants and young children. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150072.

Over, H., & Carpenter, M. (2009). Priming third-party ostracism increases affiliative imitation in children. *Developmental Science*, *12*(3), F1–F8.

Over, H., & Carpenter, M. (2013). The social side of imitation. *Child Development Perspectives*, *7*(1), 6–11.

Pagel, B., Heed, T., & Röder, B. (2009). Change of reference frame for tactile localization during child development. *Developmental Science*, *12*(6), 929–937.

Piaget, J. (2013). *Play, dreams and imitation in childhood*. Routledge.

Powell, L. J., & Spelke, E. S. (2018). Human infants' understanding of social imitation: Inferences of affiliation from third party observations. *Cognition*, *170*, 31–48.

Rakoczy, H., Hamann, K., Warneken, F., & Tomasello, M. (2010). Bigger knows better: Young children selectively learn rule games from adults rather than from peers. *British Journal of Developmental Psychology*, *28*(4), 785–798.

Ray, E., & Heyes, C. (2011). Imitation in infancy: The wealth of the stimulus: Imitation in infancy. *Developmental Science*, *14*(1), 92–105. https://doi.org/10.1111/j.1467-7687.2010.00961.x

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, *26*(6), 867–891.

Rothi, L. J., Ochipa, C., & Heilman, K. M. (1991). A cognitive neuropsychological model of limb praxis. *Cognitive Neuropsychology*, *8*(6), 443–458.

Rumiati, R., & Tessari, A. (2002). Imitation of novel and well-known actions. *Experimental Brain Research*, *142*(3), 425–433. https://doi.org/10.1007/s00221-001-0956-x

Schmidt, R. C., & Richardson, M. J. (2008). Dynamics of interpersonal coordination. In *Coordination: Neural, behavioral and social dynamics* (pp. 281–308). Springer.

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime reference guide*. Psychology Software Tools, Inc.

Schofield, W. N. (1976). Do children find movements which cross the body midline difficult? *Quarterly Journal of Experimental Psychology*, *28*(4), 571–582. https://doi.org/10.1080/14640747608400584

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

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge University Press.

Sirigu, A., Grafman, J., Bressler, K., & Suderland, T. (1991). Multiple representations contribute to body knowledge processing evidence from a case of autotopagnosia. *Brain*, *114*(1), 629–642. https://doi.org/10.1093/brain/114.1.629

Simpson, E. A., Murray, L., Paukner, A., & Ferrari, P. F. (2014). The mirror neuron system as revealed through neonatal imitation: Presence from birth, predictive power and evidence of plasticity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1644), 20130289.

Srinivasan, S. M., Lynch, K. A., Bubela, D. J., Gifford, T. D., & Bhat, A. N. (2013). Effect of interactions between a child and a robot on the imitation and praxis performance of typically developing children and a child with autism: A preliminary study. *Perceptual and Motor Skills*, *116*(3), 885–904.

Taylor, M., Cartwright, B. S., & Bowden, T. (1991). Perspective taking and theory of mind: Do children predict interpretive diversity as a function of differences in observers' knowledge? *Child Development*, *62*(6), 1334–1351.

Temprado, J. J., & Laurent, M. (2004). Attentional load associated with performing and stabilizing a between-persons coordination of rhythmic limb movements. *Acta Psychologica*, *115*(1), 1–16.

Tessari, A., Bosanac, D., & Rumiati, R. I. (2006). Effect of learning on imitation of new actions: Implications for a memory model. *Experimental Brain Research*, *173*(3), 507–513.

Tessari, A., Canessa, N., Ukmar, M., & Rumiati, R. I. (2007). Neuropsychological evidence for a strategic control of multiple routes in imitation. *Brain*, *130*(4), 1111–1126.

Tessari, A., & Cubelli, R. (2014). Route selection in action imitation: A matter of strategic choice? *Cortex*, *57*, 277–278. https://doi.org/10.1016/j.cortex.2014.01.014

Tessari, A., & Ottoboni, G. (2022). Does the body talk to the body? The relationship between different body representations while observing others' body parts. *British Journal of Psychology*, *113*(3), 758–776. https://doi.org/10.1111/bjop.12558

Tessari, A., Ottoboni, G., Symes, E., & Cubelli, R. (2010). Hand processing depends on the implicit access to a spatially and bio-mechanically organised structural description of the body. *Neuropsychologia*, *48*(3), 681–688.

Tessari, A., Proietti, R., & Rumiati, R. I. (2021). Bottom-up and top-down modulation of route selection in imitation. *Cognitive Neuropsychology*, *38*(7–8), 515–530.

Tessari, A., & Rumiati, R. I. (2002). Motor distal component and pragmatic representation of objects. *Cognitive Brain Research*, *14*(2), 218–227.

Tessari, A., & Rumiati, R. I. (2004). The strategic control of multiple routes in imitation of actions. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(6), 1107–1116. https://doi.org/10.1037/0096-1523.30.6.1107

Tikhanoff, V., Fitzpatrick, P., Nori, F., Natale, L., Metta, G., & Cangelosi, A. (2008). The icub humanoid robot simulator. *IROS Workshop on Robot Simulators*, *22*, 1–2.

Tomasello, M. (2003). *The cultural origins of human cognition (4. Print)*. Harvard University Press.

Toraldo, A., Reverberi, C., & Rumiati, R. I. (2001). Critical dimensions affecting imitation performance of patients with ideomotor apraxia. *Cortex*, *37*(5), 737–740. https://doi.org/10.1016/S0010-9452(08)70628-2

Toraldo, A., & Shallice, T. (2004). Error analysis at the level of single moves in block design. *Cognitive Neuropsychology*, *21*(6), 645–659.

Tunçgenç, B., & Cohen, E. (2016). Movement synchrony forges social bonds across group divides. *Frontiers in Psychology*, *7*, 782.

Ullman, H., Almeida, R., & Klingberg, T. (2014). Structural maturation and brain activity predict future working memory capacity during childhood development. *Journal of Neuroscience*, *34*(5), 1592–1598.

Want, S. C., & Harris, P. L. (2002). How do children ape? Applying concepts from the study of non-human primates to the developmental study of 'imitation' in children. *Developmental Science*, *5*(1), 1–14. https://doi.org/10.1111/1467-7687.00194

Wapner, S., & Cirillo, L. (1968). Imitation of a Model's hand movements: Age changes in transposition of left-right relations. *Child Development*, *39*(3), 887–894. https://doi.org/10.2307/1126991

Whiten, A., McGuigan, N., Marshall-Pescini, S., & Hopper, L. M. (2009). Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2417–2428. https://doi.org/10.1098/rstb.2009.0069
Wilson, J. L., Scott, J. H., & Power, K. G. (1987). Developmental differences in the span of visual memory for pattern. *British Journal of Developmental Psychology*, *5*(3), 249–255.
Wohlschläger, A., Gattis, M., & Bekkering, H. (2003). Action generation and action perception in imitation: An instance of the ideomotor principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1431), 501–515. https://doi.org/10.1098/rstb.2002.1257
Wood, L. A., Kendal, R. L., & Flynn, E. G. (2012). Context-dependent model-based biases in cultural transmission: Children's imitation is affected by model age over model knowledge state. *Evolution and Human Behavior*, *33*(4), 387–394.
Zentall, T. R. (2012). Perspectives on observational learning in animals. *Journal of Comparative Psychology*, *126*(2), 114–128. https://doi.org/10.1037/a0025381
Zheng, Z., Das, S., Young, E. M., Swanson, A., Warren, Z., & Sarkar, N. (2014). Autonomous robot-mediated imitation learning for children with autism. In *2014 IEEE international conference on robotics and automation (ICRA)* (pp. 2707–2712). IEEE.
Zmyj, N., & Seehagen, S. (2013). The role of a model's age for young children's imitation: A research review. *Infant and Child Development*, *22*(6), 622–641.

How to cite this article: Ottoboni, G., Toraldo, A., Proietti, R., Cangelosi, A., & Tessari, A. (2023). Paradoxical decrease of imitation performance with age in children. *British Journal of Psychology*, *00*, 1–32. https://doi.org/10.1111/bjop.12644

APPENDIX A

FINDING THE AGE AT WHICH SCORES STABILIZE (OR BEGIN TO RISE)

Identifying the age at which error rate (ER) reaches a minimum or stabilization level is not a trivial task. Pairwise statistical comparisons between consecutive age groups, searching for a non-significant difference that would correspond to the stabilization point, is little reliable (as dedicated Monte-Carlo simulations showed). Indeed, even if there is no flattening at all in the studied age range, but a weak,

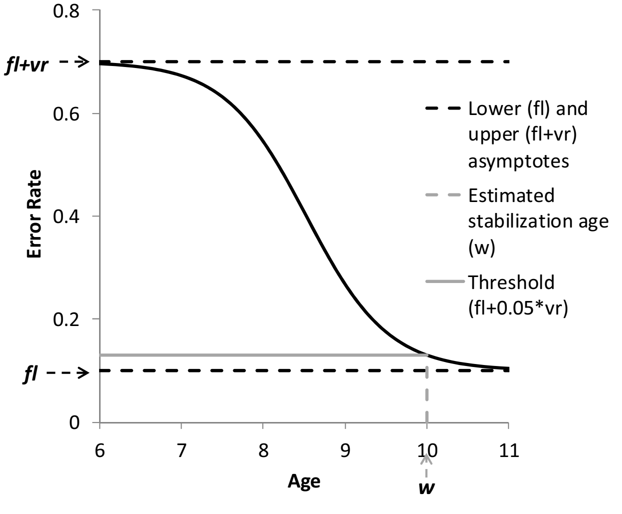

FIGURE A1 The modified logistic law (black solid curve) used to estimate the Age (*w*) at which error rate stabilizes. The logistic law is obtained by non-linear regression on empirical data. Dashed black lines: upper asymptote (floor, parameter *fl*) and lower asymptote (ceiling, sum of parameters *fl* and *vr*). Parameter *vr* (vertical range) is the distance between the asymptotes. The grey solid line (Threshold) is the error rate that is 5% of the way between the lower and the upper asymptote, and which identifies the Age were the curve (conventionally) flattens (parameter *w*, dashed grey segment). If the curve increases, the same geometry leads to the estimation of the conventional point where the curve begins to rise.

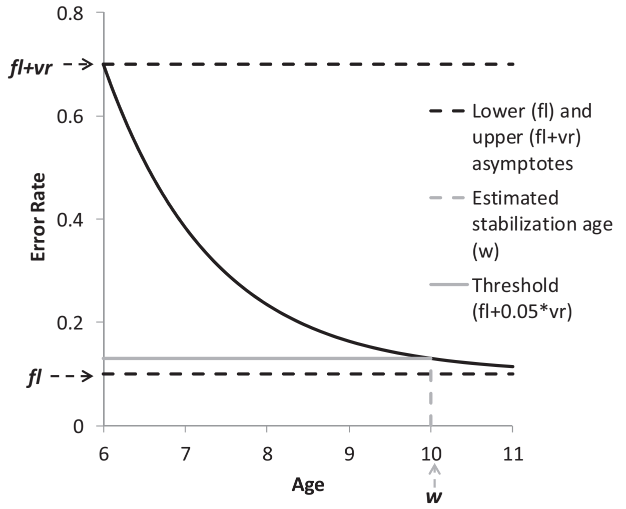

FIGURE A2 The modified exponential decay law. See Figure A1 for conventions.

progressive decrease of ER, the probability of finding a false negative (a non-significant contrast between two consecutive age groups) is very high, and this would occur anywhere along the Age scale. Thus, we reasoned that a reliable technique should not be based on pairwise comparisons but rather on accurately estimating the whole curve linking Age to ER. Given that ERs are on a bounded 0–1 scale, effects by Age cannot generally be linear. A standard way to model such effects is to fit a sigmoidal curve. The obvious candidate would be a logistic function, with an inflexion point at ER = 0.5, a lower asymptote at ER = 0 and an upper asymptote at ER = 1. However, ER certainly has a physiological above-zero lower limit. The upper limit cannot be safely assumed to be 1 either, because there might be actions for which errors are virtually impossible, especially in some categories. Thus the 0 lower asymptote was replaced by a parameter *fl* ('floor'), and the 1 upper asymptote was replaced, for mathematical convenience, with *fl* + vr, where *vr* is the vertical range (the vertical distance between the upper and the lower asymptotes). Hence, the classical logistic model, $\text{ER} = 1/(1+\exp(-(s\,\text{Age}+i)))$, with s = slope and i = intercept, became:

$$\text{ER} = fl + vr/(1+\exp(-(s\text{Age}+i)))$$

However, we are not really interested in slope ad intercept *per se*: the purpose of the present analyses is to understand at what age ER flattens after an initial drop (note that ER can also rise, and in this case, we are interested in the Age at which ERs begins to rise; the geometry is perfectly left–right symmetrical with respect to the ER-drop case, so we focused on Age-related drops in the following explanation). Clearly, the drop reaches the lower asymptote only at infinity, so one must set an ER level that is very close to the lower asymptote and consider the remaining part of the curve as flat. How close is 'very close?' We conventionally set the threshold at 5% of the overall distance (*vr*) from the lower to the upper asymptote. Thus, the threshold-ER was $T = fl + 0.05vr$. For instance (Figure A1), if the lower asymptote was estimated to be $fl = 0.1$ and the upper asymptote to be 0.7 ($vr = 0.7 - 0.1 = 0.6$), stabilization was considered to have been reached at the Age w where ER equals the threshold $T = 0.1 + 0.05$ x0.6 = 0.13.

So, we had to reparametrize the Equation so as to have w—the Age value where the curve intersects threshold T—as a parameter. Reparameterization led to the replacement of intercept i with $-\ln((1-0.05)/0.05) - ws$; hence the final regression equation became:

$$\text{ER} = fl + vr/(1+\exp(-(s\text{Age} - \ln((1-0.05)/0.05) - ws)))$$

Figure A1 shows the curve with parameters $fl = 0.1$, $vr = 0.6$, $s = -2$ and $w = 10$.

TABLE A1 Parameter estimates from non-linear regression analyses.

	Parameters			
Y variable	Floor (*fl*)	Vertical range (*vr*)	Slope (*s*)	Stabilization age (*w*)
Overall ER[a]	0.453 ±0.035	0.129 ±0.051		8.97 ±3.24
Side ER	0.068 ±0.027	0.195 ±0.05	2.83 ±2.15	7.72 ±0.89[b]
Proximal ER	0.044 ±0.023	0.158 ±0.038	−1.81 ±1.06	9.75 ±1.02

Note: ER, error rate. Mean ± Standard Error are provided for each parameter estimate. For Side and Proximal ERs, the modified Logistic Regression model reached convergence on stable parameter estimates.

[a]For Overall ER, convergence failure of the Logistic model led to the use of the Exponential Decay model, from which parameter estimates are reported.

[b]Side ER increased with age (slope *s* is positive), so the *w* parameter expresses the age at which ER began to rise.

Sometimes this modified logistic model (which was fitted with non-linear regression in R, *nls* function) failed to converge. In those cases, we replaced it with a modified exponential decay model (only useful for ER *dropping* with Age). Here the general formula was ER = exp($-s$ Age). Since the top score (ER = 1) can only be reached at Age = 0, we replaced Age with Age − 6 (the youngest participants were 6-year-old children), so that the equation could fit the full range of ERs. Again we implemented a lower physiological limit that can be higher than zero, *fl*, and a vertical range *vr*, as free parameters:

$$\mathrm{ER} = fl + vr \times \exp(-s(\mathrm{Age} - 6))$$

Reparametrization, with Age − 6 value w at which ER reaches threshold T, led to $s = -\ln(0.05)/w$; hence the final equation:

$$\mathrm{ER} = fl + vr \times \exp((\mathrm{Age} - 6)\ \ln(0.05)/w)$$

An example is shown in Figure A2, plotting the curve with parameters $fl = 0.1$, $vr = 0.6$ and $w = 4$ (corresponding to Age = 10).

Non-linear regressions were run on the per-subject ERs (mean across the three Conditions). Table A1 reports the parameter estimates that were obtained from all such analyses.

APPENDIX B

ESTIMATION OF PER-LIMB ERROR RATES: THE INVERSE TRANSFORMATION

At the time of statistical analyses, the data from each single limb on each Double-effector trial were not available anymore; however, we could estimate the per-limb ER in Double-effector conditions by a mathematical transformation that was derived theoretically and validated by a dedicated Monte-Carlo simulation study as well as by empirical evidence.

The inverse transformation

As an example, consider a test in which young children have to copy two digits. If the probability of making a mistake when copying a digit is X, and the performance on one digit does not influence the performance on the other, then the overall probability of failing the two-digit test (i.e. of miscopying the first digit, or the second, or both) is $Y = 2X - X^2$. Thus if one has an empirical estimate of Y, that is, the raw ER on the test, the per-digit error probability X can be estimated by applying the inverse transformation, $X = 1 - (1 - Y)^{1/2}$, to the raw ER. Out of the example, an unbiased estimate of the per-limb error probability X in a Double-effector condition can be obtained by applying the inverse transformation to the raw ER of that condition, that is, by computing $1 - (1 - \mathrm{ER})^{1/2}$. This unbiasedness holds under the assumptions that in the Double-effector condition, errors on arm and leg have identical probabilities and are statistically independent. By means of Monte-Carlo simulations, we studied the consequences of violations of either assumption.

Assumption of equal rates of arm and leg errors in Double-effector trials

The $1-(1-ER)^{1/2}$ transformation is robust with respect to violations of the assumption of equal ERs for arm and leg. A dedicated Monte-Carlo simulation showed that with a 0.26 (huge) difference in baseline error probability for arm and leg (0.05 vs. 0.31), the bias in the estimation of the per-limb (average) ER was only 0.01. Moreover, arm and leg, when compared with one another in the Single-effector condition, had very similar ERs. Indeed, when running GMLM on each of the three error categories that underwent the inverse transformation, the between-limbs differences were never significant: Side errors, Wald $\chi^2 = 2.272$, $p = .132$ (Arm = 0.086, Leg = 0.108); Proximal errors, Wald $\chi^2 = 0.033$, $p = .856$ (Arm = 0.045, Leg = 0.039); Final Position errors, Wald $\chi^2 = 0.373$, $p = .542$ (Arm = 0.013, Leg = 0.001).

Assumption of independent error probabilities for arm and leg

The inverse transformation is less robust with respect to violations of the assumption that errors on arm and leg in a Double-effector condition are statistically independent. Monte-Carlo simulations showed that, if errors on arms and legs are positively correlated, the per-limb ER will be underestimated by the inverse transformation; if errors on the two limbs are negatively correlated, an overestimation will occur. Luckily, these biases are small across the range of Double-effector ERs that were empirically observed across all ages and error categories (the top ER, 0.352, was recorded for Side errors by 10 year olds in the Heterolateral condition): for simulated correlations ranging from −0.5 to +0.5, biases ranged from −0.061 and +0.018.

Simulations of plausible cognitive scenarios

We wished to understand whether the above biases in the estimation of per-limb ERs in Double-effector trials contributed to the empirical results described in our work. We proceeded as follows. We selected the most frequent error category, Side errors, for which possible biases would be maximal (biases are proportional to ERs); we then simulated the cognitive processes/limits that were hypothesized to be involved in the genesis of Side errors and produced a set of predicted *raw*-ER patterns; as a last step, we compared these

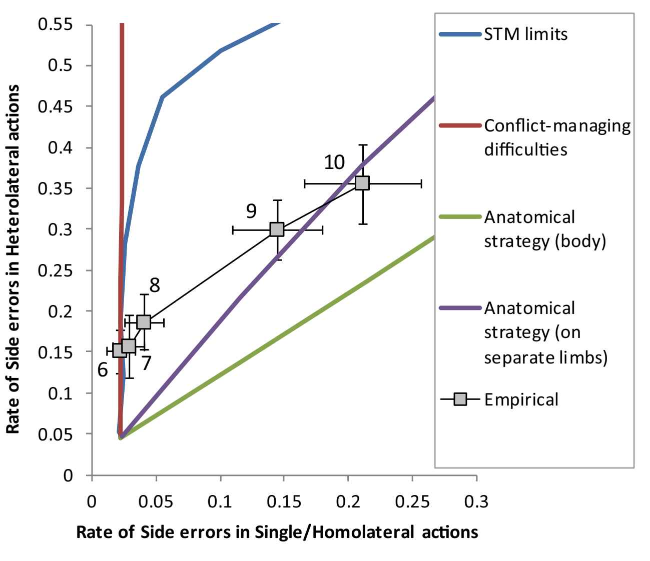

FIGURE B1 Comparison of empirical data on the raw rates of Side errors (grey squares, mean ± SE) with the predictions of four theoretical scenarios (coloured curves). A lower bound for the Single-effector error rate was set at 0.0225 (the lowest empirical finding in our experiment). See text for details.

predicted patterns with those that were observed empirically. If a strong discrepancy were found, then our hypotheses, which were derived on the grounds of the (biased) per-limb ERs, would be falsified, and this would be indirect evidence that the biases were probably rather large and might have misled our search for explanations. Else, if the observed raw-ER patterns showed good agreement with the predicted ones, then our cognitive hypotheses would be confirmed, and this would provide indirect evidence that the biases, if present, were small enough not to lead us to the formulation of wrong explanations.

Monte-Carlo simulations (*N* between 110,000 and 130,000 data points) were run for four cognitive processes/limits we proposed as explanations of the results observed on Single-effector and Heterolateral trials. The cognitive processes/limits were: (i) a limit in STM for the side(s) of the limb(s); (ii) an inability to manage the conflict between the side of one limb and the side of the other limb; (iii) an anatomical imitation strategy based on the whole body; (iv) an anatomical imitation strategy applied independently to the two limbs. Note that (i), (ii) and (iii) entail violations in the assumption of independence of errors between arm and leg in Heterolateral trials, which predict biases in the estimation of per-limb ERs. (i) Because of STM limitations, forgetting the side of one limb is much more frequent than forgetting the sides of both limbs; so, if the side of one limb is forgotten, the side of the other is more likely remembered than not. This causes a negative correlation between errors on the two limbs, which in turn yields overestimation of the per-limb ER. (ii) In the 'conflict-managing limit' scenario, the side of one limb is mistaken as the side of the other limb (e.g. left arm and right leg become left arm and left leg) because the child is not able to manage the conflict between the spatial labels ('left' vs. 'right'). The extra-errors due to this phenomenon would all (or mostly) regard a single limb per trial; thus, the predicted pattern is qualitatively similar to that implied by the STM scenario, with a negative correlation between errors on the two limbs and final overestimation of the per-limb ER. (iii) The hypothesis of an anatomical imitation strategy in which the participant reverses the whole body's laterality codes, predicts that Side errors would either regard both limbs simultaneously, or neither of them. This would cause a strong positive correlation between errors on the two limbs, which in turn would yield underestimation of the per-limb ER. (iv) If the anatomical imitation strategy were not applied to the whole body, but separately and in a statistically independent way, to the arm and the leg, Side errors on both limbs would be uncorrelated, so, the inverse transformation would yield an unbiased estimation of the per-limb Heterolateral ER (which would, notably, be identical to the ER on Single-effector actions).

The general explanation we proposed for the results of our experiment specified that younger children are characterized by a STM/conflict-managing limit, without anatomical imitation strategy, while older children show exactly the opposite profile. If such an explanation is correct, the empirical data should progressively 'shift' from the pattern predicted by the STM/conflict-managing-limit model (6 year olds) to the pattern predicted by the anatomical imitation-strategy model (10 year olds).

Figure B1 shows the comparison between predictions and empirical data. It plots the raw rate of Side errors in Single/Homolateral actions (X) against that in Heterolateral actions (Y). The coloured curves (obtained from the Monte-Carlo simulations) show the predicted patterns on grounds of different degrees of the four types of cognitive limitations: (i) STM limits (blue), (ii) conflict-managing difficulties (red), (iii) anatomical imitation strategy applied to the whole body (green) and (iv) anatomical imitation strategy applied to each limb separately (violet). Different points within each curve represent different degrees in the simulated phenomenon; thus, as one 'moves' downwards over the blue curve, STM span increases; moving upwards along the red line, the inability to manage conflicts worsens; moving upwards along the violet or green lines, the probability that an anatomical imitation strategy is applied, increases (note that a 0.0225 baseline ER was implemented for both limbs, over and above all the simulated processes; it corresponded to the minimum ER recorded across all age groups and conditions).

Empirical data (grey squares) are shown for different ages (6–10). Consistently with our explanation, younger children's performance was compatible with the hypotheses of STM and/or conflict-managing difficulties (red and blue curves), and older children's performance was compatible with the hypothesis of an anatomical imitation strategy. Thus, following the previously mentioned reasoning, this is indirect evidence that the biases in the estimation of per-limb ERs were likely small: had they been large, they would have led us to hypotheses whose predictions on raw ERs would have failed to fit the observed data. This did not happen.

Interestingly, the prediction with which the results of 10 year olds were most compatible, was not that of an anatomical imitation strategy applied to the whole body (green curve), but rather, that of an anatomical imitation strategy applied *separately* to the two limbs (violet curve). Is it at all possible to apply an anatomical imitation strategy to the arm and not to the leg (or vice versa) within the same trial? Clearly, the notion of analogical mental rotation clashes with this idea. One possible assumption might be that anatomical imitation is not carried out by internal rotation, but rather, by propositionally changing some abstract, spatial labels, possibly separate for arm and leg. However, such reasoning is definitely premature. We cannot exclude that older children's performance had *both* relevant STM limits (blue curve) and a body-based anatomical imitation strategy (green curve)—their performance would have fallen somewhere in between the two curves, as it in fact happened, and the fact that it fell over the violet curve would be just a coincidence. To the best of our knowledge, there is no way to tell which interpretation is correct, the one assuming an anatomical imitation strategy applied separately to the two limbs (iv, violet), or the one assuming a combination of STM limits (i, blue) and body-based anatomical imitation (iii, green), at least, not on the grounds of the present experiment: future research will address this interesting issue.

APPENDIX C

DESCRIPTION OF THE STIMULUS ACTIONS

See Table C1.

TABLE C1 Description of the 12 actions employed in this study.

Condition	Limb(s)	ID	Description	Side(s)
Single	Arm	1	The model lifted the arm laterally, keeping it extended, with abduction of the shoulder, until the hand reached the height of the shoulder. At that point, the hand rotated by 90°, bringing the palm to face forwards, then the movement was undone. The arm was then brought back to starting position	L or R
Single	Arm	2	The model lifted the arm forward, keeping it extended, with flexion of the shoulder, until the hand reached the height of the shoulder. At the end of the movement, the palm faced down. After that, the hand was rotated by 135°, till the palm faced medially/upwards, then the movement was undone. The arm was finally brought back to starting position	L or R
Single	Leg	3	The model lifted the leg forward (flexing the hip) and flexed the knee to keep the calf at rest. The lifting movement stopped when the hip had a 90° angle (i.e. when the thigh was parallel to the ground). At that point, the knee was rotated medially (inward), lifting the foot, by 45° (Figure 1a), then the movement was undone. The leg then went back to starting position	L or R
Single	Leg	4	The model lifted the leg forward (flexing the hip), keeping it extended, until it formed a 45° angle with the ground. Then the foot-tip was rotated medially by 45°, then laterally by the same amount (foot-tip like an upward pendulum) and back to starting position	L or R
Homolateral	Arm+Leg	5	Simultaneous movements of arm and leg of the same body side Arm: the model lifted the arm forward, keeping it extended, with flexion of the shoulder, until the hand (palm facing medially) reached the height of the shoulder Leg: the model lifted the leg forward (flexing the hip) and flexed the knee to keep the calf at rest. The lifting movement stopped when the thigh was parallel to the ground Both limbs were then (simultaneously) brought to starting position	LL or RR

(Continues)

TABLE C1 (Continued)

Condition	Limb(s)	ID	Description	Side(s)
Homolateral	Arm+Leg	6	Simultaneous movements of arm and leg of the same body side (Figure 1b) Arm: the model lifted the arm laterally, keeping it extended, with abduction of the shoulder, until the arm pointed upwards, over the head (a full, 180° rotation of the arm). At the end of the movement, the hand-palm faced outwards Leg: the model lifted the leg forward (flexing the hip) and flexed the knee to keep the calf at rest. The lifting movement stopped when the thigh was parallel to the ground Both limbs were then (simultaneously) brought to starting position	LL or RR
Homolateral	Arm+Leg	7	Simultaneous movements of arm and leg of the same body side Arm: the model lifted the arm forward, keeping it extended, with flexion of the shoulder, until the hand (palm facing medially) reached the height of the shoulder Leg: the model lifted the leg laterally (abducting the hip), keeping it extended, until it formed a 45° angle with the other leg Both limbs were then (simultaneously) brought to starting position	LL or RR
Homolateral	Arm+Leg	8	Simultaneous movements of arm and leg of the same body side Arm: the model lifted the arm laterally, keeping it extended, with abduction of the shoulder, until the arm pointed upwards, over the head (a full, 180° rotation of the arm). At the end of the movement, the hand-palm faced outwards Leg: the model lifted the leg laterally (abducting the hip), keeping it extended, until it formed a 45° angle with the other leg Both limbs were then (simultaneously) brought to starting position	LL or RR
Heterolateral	Arm+Leg	9	As in 5, but arm and leg were of opposite body sides	LR or RL
Heterolateral	Arm+Leg	10	As in 6, but arm and leg were of opposite body sides	LR or RL
Heterolateral	Arm+Leg	11	As in 7, but arm and leg were of opposite body sides (Figure 1c)	LR or RL
Heterolateral	Arm+Leg	12	As in 8, but arm and leg were of opposite body sides	LR or RL

Note: In his/it's starting position, the model stood upright, with slightly separate legs and the arms in resting position along with the body, with the hand-palms facing the thighs and slightly flexed fingers.

'Condition' reports the three levels of action complexity: Single (actions involving only one arm or one leg), Homolateral (actions involving one arm and one leg of the same body side), Heterolateral (actions involving one arm and one leg of opposite body sides). 'Limb(s)' reports whether the arm, the leg, or both were involved in the action; 'Side(s)' specifies the side(s) of the limb(s) involved.