Journal of Experimental Child Psychology 239 (2024) 105826

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Imitation fidelity increases with age in boys, but not in girls: An intriguing finding in a cohort of children aged 3 to 6 years

Natàlia Barbarroja [a], Mikhail Ordin [b], Michael J. Walker [c], Sandra Posada-Salazar [d], Héctor M. Manrique [e,*]

[a] *Department of Linguistics and Basque Studies, University of Basque Country–UPV/EHU, 01006 Vitoria-Gasteiz, Spain*
[b] *Coimbra Institute for Biomedical Imaging and Translational Research, Institute for Nuclear Sciences Applied to Health, University of Coimbra, 3000-548 Coimbra, Portugal*
[c] *Department of Zoology and Physical Anthropology, Faculty of Biology, University of Murcia, 30100 Murcia, Spain*
[d] *Independent Researcher, Barcelona, Spain*
[e] *Department of Psychology and Sociology, Universidad de Zaragoza, Campus Universitario de Teruel, 44003 Teruel, Spain*

A R T I C L E   I N F O

Article history:
Received 6 September 2023
Revised 8 November 2023
Accepted 9 November 2023
Available online 19 December 2023

Keywords:
Children
Instrumental task
Imitation
Emulation
Normativity
Boy–girl differences

A B S T R A C T

Imitation that entails faithful reproduction of demonstrated behavior by reenacting a sequence of actions accurately is a fast and efficient way to acquire new skills as well as to conform to social norms. Previous studies reported that both culture and gender might impinge on young children's fidelity of imitation. We analyzed the imitative behavior of 87 children whose ages ranged from 3 to 6 years. An instrumental task was administered that offered partial (opaque apparatus) or total (transparent apparatus) information about causal connection between the demonstrated actions and their effect in achieving a desired reward. Imitative fidelity (imitating the actions that were demonstrated by an adult model yet were unnecessary for achieving the instrumental goal) increased as a function of age in boys, whereas no differences were found in girls. This lack of increase in girls can be ascribed to their displaying higher degrees of imitation fidelity at an earlier age.

© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

* Corresponding author.
E-mail address: manrique@unizar.es (H.M. Manrique).

https://doi.org/10.1016/j.jecp.2023.105826
0022-0965/© 2023 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

Social learning encompasses different types of knowledge acquisition through social contact (Heyes, 2012), which extend from learning that interesting things occur at a specific "spot" (e.g., stimulus enhancement), to faithful, orderly reproduction of observed actions of others (Byrne & Russon, 1998). Authors often differentiate between emulation and imitation (with "over-imitation" as a particular case of the latter). By emulation, we understand the reproduction of demonstrated goals or outcomes, such that observers "reproduce the completed goal ... by whatever means" (Tomasello, 1990, p. 284). Imitation ("true" imitation), on the other hand, entails faithful reproduction of demonstrated behavior by reenacting a sequence of actions precisely and accurately (cf. "impersonation"; Byrne & Russon, 1998, p. 675). Over-imitation is the "imitation of perceivably causally unnecessary actions in relation to the goal of an action sequence performed by a model" (Hoehl et al., 2019, p. 91). Over-imitation could be explained by an "automatic causal encoding" (Lyons et al., 2011); that is, seemingly irrelevant actions performed intentionally by a model (demonstrator) are encoded automatically as supposedly necessary and hence "causally relevant" (Lyons et al., 2007, 2011). As children become more aware of what is causally connected to task resolution, they can relegate superfluous actions by a "copy and correct" process (Whiten et al., 2005, 2009).

Another interpretation is that children associate themselves with socially desirable models who are imitated closely; imitation of unnecessary actions may signal "identification" or reflect a desire to affiliate with the model (Hobson & Hobson, 2008; Nielsen & Blank, 2011; Over & Carpenter, 2012; Tomasello et al., 2005), in support of which Nielsen (2006) found that 24-month-old children imitated actions performed by a model acting socially, although not if acting nonsocially. In a like vein, imitation of irrelevant actions by children is higher when in the presence of an inefficient demonstrator, yet lower in the presence of an efficient one, thereby suggesting that imitation in this context might serve an affiliative role rather than an exploitative one (Nielsen & Blank, 2011). Alternatively, children may regard models as "normative" who therefore should be imitated; conformity here underpins reproduction of unnecessary actions should children feel that they need to imitate all the actions performed by a model because they assign normative force to these (Kenward et al., 2011; Keupp et al., 2013). Normativity in children has been researched extensively, and its developmental pattern is well-defined; for instance, at 2 years of age children already follow adult requests and exhibit basic social conformity (Rakoczy & Schmidt, 2013), and at 3 years of age they spontaneously regard behavior by adults as normative without requiring explicit instruction from them (Vaish et al., 2011) and may protest by using normative language (e.g., "That's wrong!") on witnessing others break rules. Three-year-olds (as opposed to 2-year-olds) understand that norms are not absolute and complain whenever they are violated in appropriate contexts (Rakoczy et al., 2009). A milestone of normative development occurs around 5 years of age, when children begin making their own rules to regulate their interactions and forestall conflict (Grueneisen & Tomasello, 2017). Normativity has been linked explicitly and experimentally to over-imitation (Keupp et al., 2013); plausibly, both follow a similar developmental pattern given that in a sample of 5- to 8-year-old children it was the oldest children who imitated more irrelevant actions (over-imitated) (Marsh et al., 2014).

Imitation, therefore, could foster the internalization of social norms (Essler et al., 2023), and it is possible that fidelity of imitation lends itself to cultural variation. In this vein, it has been reported that children from the ethnic group Aka, hunter-gatherers of the Congo Basin, display unusually higher levels of emulation to solve an instrumental task, and it takes developmentally longer (~7 years) for them to switch to imitating with fidelity (Berl & Hewlett, 2015). On the other hand, Malaysian 4- to 6-year-old children from a rural indigenous community showed higher fidelity of imitation than their town-dwelling Malaysian counterparts and a sample of urban Australian children while using tools to solve a puzzle box (Fong et al., 2021). Apart from the mediating effect that culture has on fidelity of imitation, biological sex might also play a significant role. A cross-cultural study with Serbian and French children found that boys belonging to both countries over-imitated more than girls while manufacturing a hook to extract a reward from a glass bottle container (Frick et al., 2017). Enhanced same-gender imitation was also reported in 3- to 5-year-old Chinese children when verbal cues that highlighted gender norm differences were used (Wang et al., 2021).

Notwithstanding the aforementioned studies, few have tackled fidelity of imitation by focusing on both biological sex and developmental age simultaneously. To fill this gap, we explored how fidelity of imitation develops in boys and girls in the age range from 3 to 6 years, which is a most significant period in the socialization of gender. Thus, whereas 2-year-old girls and boys are merely aware of their being physically different, a high proportion of 3-year-olds already understand that these morphological differences are permanent invariant features (Bem, 1989) and the first instances of segregation by sex while playing are reported to occur (Hartup, 1983), which reflect the growing impact of biology on behavior. At 6 years of age, segregation by sex is very ostensive, and so is the adoption of gender roles. Therefore, in the age period from 3 to 6 years, critical changes in the socialization of gender take place that originate in sexual biological dimorphism and later are fostered, at least in part, by families', peers', and teachers' expectations and attitudes (Grace et al., 2008; Serbin et al., 1994; Pitcher & Schultz, 1983).

Experimental considerations

We were inspired by Horner and Whiten (2005), who presented 3- and 4-year-old children and chimpanzees with an instrumental task, based on two similar apparatuses differing only in the exterior transparency of one and the exterior opacity of the other, and showed them how to extract a reward by demonstrating both necessary (causally relevant) and unnecessary actions. Children, unlike chimpanzees, reproduced unnecessary actions performed by the model in not only the opaque apparatus but also the transparent one where the reward was attainable without them.

According to social interpretations of over-imitation, the copying of unnecessary or superfluous actions reflects a normative force that is an endorsement of the behavior of the demonstrator (Keupp et al., 2013). Because important milestones in normative development occur across the age period investigated here (3–6 years), we expected to find a steady increase in imitation fidelity as a function of age. Nevertheless, we might have also expected to find differences by sex in imitation accuracy because it has been proposed that imitation accuracy fulfills an affiliative role (Hobson & Hobson, 2008; Nielsen & Blank, 2011; Over & Carpenter, 2012; Tomasello et al., 2005) and it is claimed that boys and girls differ in their affiliative drive (Bakan, 1966; Maccoby & Jacklin, 1974), although the magnitude and direction of these differences are unclear (Bakan, 1966; Benenson et al., 2012; Duncan & Peterson, 2010). To investigate whether these hypotheses might be validated empirically, we built analogous equipment (see Fig. 2 in "Materials" section) to that deployed by Horner and Whiten (2005). The degree of accuracy to which a demonstrator's behavior is reproduced perhaps might depend on one or more of the following factors: (a) children's desire to affiliate with demonstrators through faithful copying of what they do, (b) children's attribution of normativity to demonstrators' behavior, and (c) causal information gauged from observing how the apparatuses worked that differs as a function of transparency of the apparatus. Furthermore, testing 3- to 6-year-old children enabled us to inquire whether imitative tendencies increased alongside developmental maturation at an age where important changes in the socialization of gender occur.

Method

Participants

We recruited a total of 93 3- to 6-year-old children (M_{age} = 4.52 years, SD = 0.8; 44 girls and 49 boys) attending three schools at Sabadell in Catalonia (*Escola Catalunya, Escola Enric Casassas,* and *Escola Ribatallada*). We needed to exclude 6 children from our analyses for the following reasons: For 4 of them we failed to obtain their dates of birth; 1 was lost on account of a technical failure caused by the camera turning off and on during the experiment; and 1 was uncooperative, making no attempt to do the task and neither moving nor speaking in response to any prompts from the experimenter (moreover, less than 1 year before our study, when this child was 3 years old, he had undergone surgery for scaphocephaly, which perhaps contributed to his social unresponsiveness). Therefore, we analyzed data from 87 children (45 boys, M_{age} in months = 58.311, median = 59, SD = 10.54, *SE* = 1.57, 95% credible interval (CI) [55.23, 61.39], normally distributed; and 42 girls, M_{age} in months = 60.6, median = 61, *SD* = 8.65, *SE* = 1.34, CI [57.98, 63.21], normally distributed). There were no observable age differences between boys and girls Student's *t* test was used to compare age in months (assumptions of normality and equality of variance were not violated) between the genders. The test yielded an insignificant result, $t(85) = 1.1$, $p = .274$. Fig. 1 shows that the distribution of age was equivalent across groups.

Our project was approved by the Ethics Committee for Medical Research at Cáceres, Spain. Parents were briefed about the study and needed to sign consents to their children's participation.

Materials

Two similar inverted T-shaped apparatuses (Fig. 2) were made from transparent and opaque materials, respectively, each consisting of a horizontally orientated tube (23.5 cm × 5.5 cm in diameter) and a vertically orientated tube (8 cm × 6 cm in diameter) united by a cable-tie, although not in open connection with each other; that is, despite being bound firmly together, because there is no hole between them, a stick placed in the vertical tube cannot enter the horizontal one. The apparatuses' exterior openings were covered with dark cloth that prevented children from seeing how the apparatuses worked and the reward within. In the opaque version, neither movement of the reward nor a putative connection stick–reward could be tracked. The transparent version, in contrast, afforded all the causal information needed to ascertain the utility of the model's actions for extracting the reward. Each apparatus was attached to a polystyrene board (42 × 30 × 0.4 cm) tied to a camera tripod by string, allowing adjustment to children's height. The reward (mostly potato crisps) lay inside a yellow glass container (5 cm × 4 cm in diameter) situated in the middle of the horizontal tube before each trial. The container was washed between sessions and apparatuses for hygienic reasons. We chose to use food as reward instead of a sticker because we felt that with such a wide age range (3–6 years) it was possible that stickers might not motivate all participants equally. After discussing this with their teachers, we decided that a potato crisp could be equally motivating across the age range. A red wooden stick (30 cm × 1 cm in diameter) was used as a probe to extract the baited container. A curtain blocked visual access while the apparatuses were being swapped across trials. The test sessions were video-recorded.

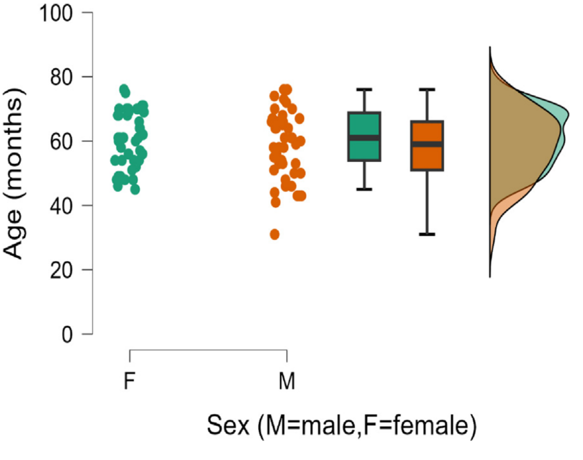

Fig. 1. Age (in months) distribution in male and female samples. Dots represent individual cases split by biological sex, box plots display the sample medians (as solid lines inside the boxes), the boxes indicate the range of 50% of the data points, and whiskers indicate the top and bottom 25% of cases, provided that they do not exceed 1.5 interquartile ranges beyond the hinges beyond the boxes. The cases that exceed this range are marked as outliers (do not present in our samples). The rightmost part of the figure displays probability density for each sample (estimated using a Gaussian kernel).

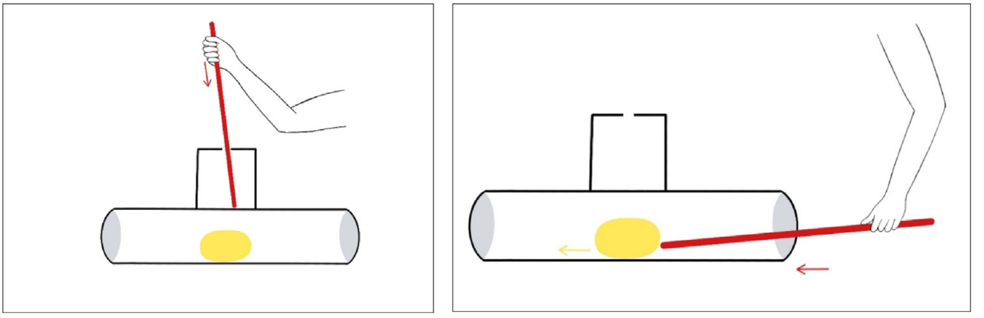

Fig. 2. Each apparatus consisted of two nonconnecting tubes, one horizontal and the other vertical, joined together in the form of an inverted T. The cavities of the two tubes were sealed off from each other because there was no hole or opening between the vertical and horizontal tubes. One apparatus was made of transparent material, and the other was made of opaque material; a small container (yellow) containing a reward was placed in the horizontal tubes. The model demonstrated actions as follows. First, a stick was inserted twice into the vertical tube (left panel), after which it was inserted into the horizontal tube (right panel), where it pushed the container to the far end. Whereas in the transparent apparatus the vertically inserted stick is clearly seen to be unable to touch the container in the horizontal tube, with the opaque apparatus its inability to do so cannot be discerned.

Procedure

Each participant was ushered into the testing room by a mentor (e.g., teacher) who gave the instruction "Sit down and wait!", followed by emphatic nonverbal gestures such as tapping the chair and extending a hand as a signal not to move. Placed 1.5 m from the child, the apparatus was adjusted to the child's line of sight. A sliding curtain initially concealed the apparatus from the child in order to reduce the likelihood of untimely interactions and to allow re-baiting and swapping across apparatuses. The order in which apparatuses were administered was fixed, always starting with the opaque apparatus; this was intentional because causal information needed to be concealed before the second demonstration took place using the transparent apparatus given that if participants had faced the transparent apparatus first, when confronting the opaque apparatus they would have already known which actions were causally related to food extraction, and hence potential comparisons would have been less informative.

After baiting the opaque apparatus, the sliding curtain was pulled back and the mentor (demonstrator/model) placed herself at 90 degrees to the apparatus (right or left according to the child's manual laterality) to perform the following demonstration. By using her right or left hand, according to the participant's manual dexterity, the demonstrator introduced the red stick into the apparatus's top hole twice and then into the horizontal tube once, thereby pushing the yellow glass container to the far end and hence extracted, whereupon its bait, a potato crisp reward, was shown, but not handed, to the child. The curtain was closed once more, and the glass container was re-baited and re-introduced into the apparatus. The curtain was then withdrawn, and the mentor commanded the child "Now you!", and with both hands open offered the red stick to the child, who was allowed 5 min of free unconstrained interaction with the apparatus. If the child managed to produce the glass container within that time, then she or he could eat the potato crisp and the trial was considered over. While the child probed the apparatus, the mentor withdrew to a corner to make herself as inconspicuous as possible. If 5 min elapsed without success, the trial was terminated. The curtain was then extended in order to swap apparatuses and prepare for testing with the transparent apparatus. Throughout the testing session, the mentor always averted the child's gaze in order to avoid giving away cues or distracting the participant from the critical stick-related actions. When testing with the opaque apparatus was completed, the exact same procedure was repeated with the transparent apparatus.

If a child had problems with the use of a potato crisp, it was replaced by something more appropriate and/or rewarding. The investigators trained children's mentors to perform the test until reliably mastering the procedure 1 week before the testing sessions began and assisted them punctually during testing in the swapping and re-baiting of apparatuses, yet they vacated the room as soon as children attempted to solve the task in order to reduce any potential audience effects.

Coding and analysis

We considered the following independent variables: (a) apparatus type (opaque vs. transparent), (b) participants' age, and (c) sex. The fidelity of imitation of each participant was scored from 0 to 7 (from least to greatest imitative accuracy). We considered both the number of steps reproduced and the order in which they were deployed. Coding was as follows: score = 0, no probing with the stick in any hole; score = 1, insertion of the stick into a single horizontal hole regardless of how often; score = 2, insertion of the stick into a single hole, the top hole, regardless of how often (the addition of 1 point here, with respect to score = 0, was because poking in the top hole gave no reward); score = 3, insertion of the stick into both holes in reverse order (first the horizontal hole and then the top hole) at least once, albeit without matching the number of insertions per hole as demonstrated; score = 4, insertion of the stick into both holes in reverse order, but with the same number of insertions per hole as demonstrated by the model (first one horizontal insertion, whereby the reward was extractable, and then probing twice in the top hole); score = 5, insertion of the stick into both holes in the correct order (first the top hole and then the horizontal hole) an incorrect number of times; score = 6, insertion of the stick into both holes in correct order and frequency, thereby reproducing the model's actions with fidelity (two insertions into the top hole and one insertion through the horizontal hole for dislodging the reward; and score = 7, after securing the reward, children showed the potato crisp and/or lifted the screen concealing the apparatus. These latter actions were not intended as demonstrations initially, yet a few children imitated them and we were forced to add one additional level score. We decided to code the fidelity of imitation across a score continuum of 0 to 7 instead of sorting what children did into two discrete categories (necessary vs. unnecessary actions) for two main reasons. First, we regarded over-imitation as a special case of true imitation where nothing "extra" was added by the imitator that was not already present in the demonstration. Second, we thought that not sorting children's actions into two dichotomous categories (necessary vs. unnecessary) might help us to capture more subtle statistical differences. Therefore, our focus was on accuracy/fidelity of imitation. N.B. was responsible for watching and coding all videos. To assess inter-observer reliability, H.M.M. coded 20% of the trials; inter-observer reliability was nearly perfect (Cohen's kappa = .95).

Results

We hypothesized that imitation increases with age. To test this hypothesis in the first approximation, we calculated correlation coefficients between age (in months) and imitation score. The correlations were run separately for male and female samples and separately for transparent and opaque apparatuses. Because the age and imitation score are ordinal variables and the range of imitation score values is limited, we ran Kendal's tau version of nonparametric correlation tests. We also undertook Bayesian analyses of the correlation tests as well as of all subsequent statistical tests, which allowed estimation of the strength of evidence for the null hypothesis. Although Kendal's tau yields lower coefficient scores than Spearman's rho, it is more accurate from the standpoint of estimating the significance of the nonparametric coefficients in the case of potential outliers, smaller sample sizes, and (most important) when the value range of one of the variables is narrower than the value range of the other variable (Gilpin, 1993), which often leads to violations of bivariate normality. Moreover, Bayesian nonparametric coefficients are based on Kendal's tau, and hence using this measure for both frequentist and Bayesian tests enhances the consistency of the analyses that stand together, forming (so to speak) a coherent "pipeline."

The hypothesis was confirmed on the male population for the transparent apparatus, $B = .319$, $p = .003$, $z = 0.331$, and for the opaque apparatus, $B = .280$, $p = .007$, $z = 0.288$ (one-tailed test because we estimated significance for the hypothesis that the correlation between age and imitation score is positive; imitation fidelity increases with age). For the Bayesian test, we used a default prior width of $\beta = 1$, which is suggested when prior knowledge is absent, vague, or difficult to elicit (Ly et al., 2016) so that all values of coefficients are equally likely as a prior. On the transparent apparatus, the Bayes factor for the correlation between imitation score and age was BF_{10} = 40.803 (the hypothesis that imitation fidelity increases with age in the male population was 40 times more likely, given the data, than the hypothesis that there is no relation between age and imitation fidelity; this is very strong evidence in favor of the relation between age and imitation fidelity). On the opaque apparatus, the Bayes factor was BF_{10} = 13.837 (the hypothesis that imitation fidelity increases with age in the male population was 14 times more likely, given the data, than the hypothesis that there is no relation between age and imitation fidelity; this is strong evidence in favor of the relation between age and imitation fidelity). These patterns are represented in Figs. 3 and 4.

We then compared whether the relation between imitation score and age was significantly stronger on the transparent apparatus than on the opaque apparatus (we compared correlation coefficients) in the male population. We treated the correlations as Pearson coefficients and used Fisher's *z*-transform for comparison, which was found to be more robust—albeit more conservative—than converting the coefficients to the Pearson equivalent prior to transformation (Myers & Sirois, 2006). The test showed that the difference between the two coefficients was not significant, z = 0.196, p = .84 (an alternative approach suggested by Walker (2003)—converting nonparametric coefficients to the Pearson's equivalent and then performing *z*-transform for comparing correlations—confirmed the conclusion that the difference in correlation strength between imitation scores and age was neither substantial nor significant).

Although imitation fidelity indeed increased with age in the male population, we did not observe any association between imitation score and age in the female population either for the transparent apparatus, B = −.004, p = .514, z = −0.004, BF_{01} = 194 (the null hypothesis was 5 times more likely than the hypothesis that imitation increases with age in the female population, which represents moderate support for the absence of association), or for the opaque apparatus, B = −.048, p = .652, z = −0.048, BF_{10} = .145 (the null hypothesis was 7 times more likely than the hypothesis that imitation increases with age in the female population, which represents moderate support for the absence of association). The difference in strength of correlations between imitation score and age on the transparent versus opaque apparatus was weak and not significant, z = 0.194, p = .85. These patterns are represented in Figs. 5 and 6.

Although we did not observe significant differences in the association strength between imitation score and age on the transparent versus opaque apparatus for either sex, there was a noticeable trend in both samples suggesting slightly higher imitation fidelity on the transparent apparatus in both

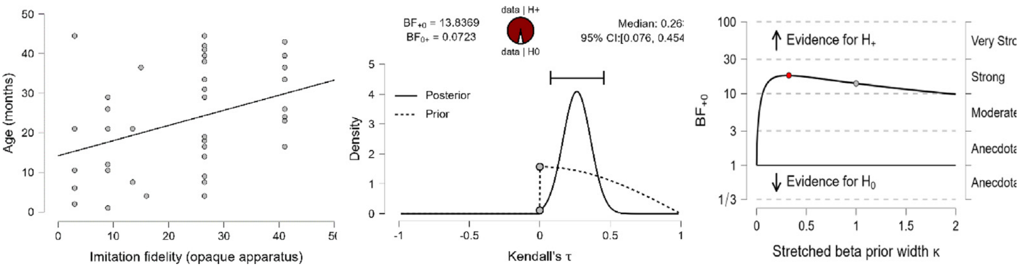

Fig. 3. Boys, opaque apparatus. Left: Correlation plot between age (in months, ranked) and imitation fidelity on opaque apparatus. The numbers on the scatterplot axes are percentage ranks used for calculating the Kendall's tau coefficient. Middle: Prior and posterior distributions of the correlation under the alternative hypothesis (H1) and the 95% credible interval (CI) of the posterior distribution (horizontal solid line). The probability wheel shows the odds of the data for the null hypothesis (H0) and alternative hypothesis under the prior that both are equally likely before the data are analyzed. The Bayes factor (BF) is the ratio between the prior and posterior at the test point, indicated by the gray circles (here, we were testing that the correlation coefficient differs from zero, which is a test point). Right: Bayes factor as a function of the width of the β prior. The curve shows possible values of the Bayes factor that could be obtained given the data if a different prior were chosen. The gray circle (right) displays the Bayes factor given the chosen β width = 1 (default prior that we selected), and the red circle (left) shows the Bayes factor that could be maximally attained given the data. Overall, the graph shows that the data provide strong evidence in favor of the alternative hypothesis irrespective of the β width selected a priori.

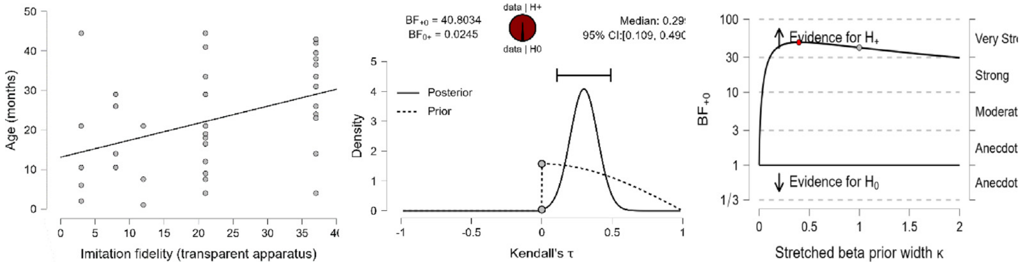

Fig. 4. Boys, transparent apparatus. Left: Correlation plot between age (in months, ranked) and imitation fidelity on transparent apparatus (in ranks). Middle: Prior and posterior distributions of the correlation under the alternative hypothesis (H1) and the 95% credible interval (CI) of the posterior distribution (horizontal solid line). Right: Bayes factor (BF) as a function of the width of the β prior. The gray circle (right) displays the Bayes factor given the chosen β width = 1 (default prior that we selected), and the red circle (left) shows the Bayes factor that could be maximally attained given the data. Overall, the graph shows that the data provide very strong evidence in favor of the alternative hypothesis irrespective of the β width selected a priori. H0, null hypothesis.

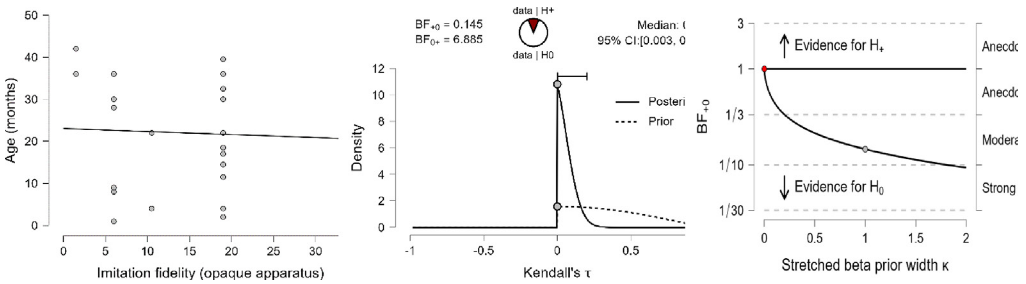

Fig. 5. Girls, opaque apparatus. Left: Correlation plot between age (in months, ranked) and imitation fidelity on opaque apparatus. The numbers on the scatterplot axes are percentage ranks used for calculating the Kendall's tau coefficient. Middle: Prior and posterior distributions of the correlation under the alternative hypothesis (H1) and the 95% credible interval (CI) of the posterior distribution (horizontal solid line). Right: Bayes factor (BF) as a function of the width of the β prior. The curve shows possible values of the Bayes factor that could be obtained given the data if a different prior were chosen. The gray circle (right) displays the Bayes factor given the chosen β width = 1 (default prior that we selected), and the red circle (left) shows the Bayes factor that could be maximally attained given the data. Overall, the graph shows that the data provide moderate evidence in favor of the null hypothesis (H0) with most prior values of the β width.

populations. Therefore, we decided to lump the data from both populations and compare imitation scores between apparatus types (due to violations of the normality assumption, we ran two-tailed Wilcoxon signed-rank tests for paired values, and the Bayesian version was run with a default prior defined by Cauchy's distribution parameter set to 0.7, centered at zero, suggesting that the null hypothesis was represented by the imitation score difference on apparatus types centered at zero). The test showed that imitation fidelity was indeed significantly higher on the transparent apparatus, with a moderate yet non-negligible effect size; however, a very small value for the Bayes factor suggests that the strength of evidence in favor of the hypothesis, given the data at hand, was anecdotal, $W = 91$, $z = -2.55$, $p = .008$, rank-biserial correlation $r_{rb} = -.55$, $BF_{10} = 2.42$, showing that the tested hypothesis was only 2.4 times more likely than the null hypothesis (i.e., the null hypothesis defines that the difference between imitation scores on the transparent and opaque apparatuses is zero). This strength of evidence calls for additional empirical evidence before we can accept as reliable the hypothesis that the imitation scores between apparatus types were different.

So far, the data suggest that higher imitation fidelity was associated with older age (the sample ranged from 3 to 6 years) in boys, but not in girls (Fig. 7 shows the raw imitation scores for boys and girls per apparatus type as a function of age). One way to account for this result could be that girls

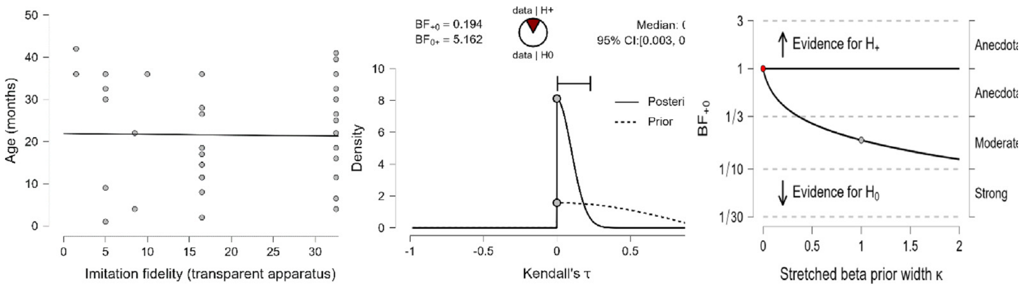

Fig. 6. Girls, transparent apparatus. Left: Correlation plot between age (in months, ranked) and imitation fidelity on transparent apparatus. The numbers on the scatterplot axes are percentage ranks used for calculating the Kendall's tau coefficient. Middle: Prior and posterior distributions of the correlation under the alternative hypothesis (H1) and the 95% credible interval (CI) of the posterior distribution (horizontal solid line). Right: Bayes factor (BF) as a function of the width of the β prior. The curve shows possible values of the Bayes factor that could be obtained given the data if a different prior were chosen. The gray circle (right) displays the Bayes factor given the chosen β width = 1 (default prior that we selected), and the red circle (left) shows the Bayes factor that could be maximally attained given the data. Overall, the graph shows that the data provide moderate evidence in favor of the null hypothesis (H0) with most prior values of the β width.

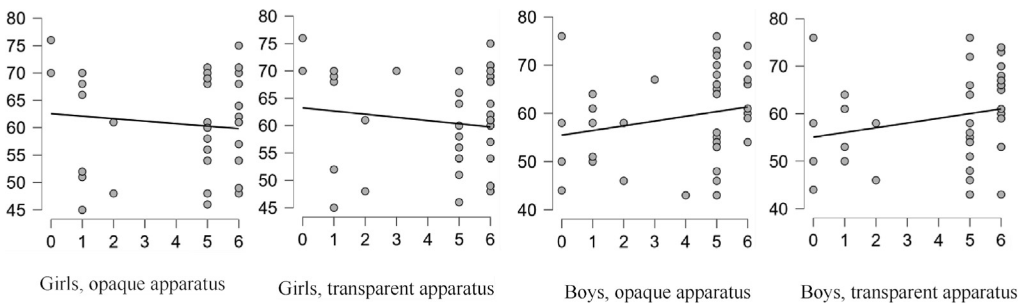

Fig. 7. Raw imitation scores as a function of age for boys and girls on transparent and opaque apparatuses. The horizontal axis represents imitation score, and the vertical axis represents age of participants (in months).

had developed imitation fidelity at a faster rate than boys, and therefore more girls reached higher imitation fidelity in this task than boys and at the group level the imitation fidelity score was higher in the female subsample. Imitation fidelity in boys was under-developed compared with that in girls in this age span. To test this hypothesis, we split our sample into younger (3- and 4-year-olds) and older (5- and 6-year-olds) groups. This split was justified because the age distribution was similar in girls and boys (see description of the sample above). If the hypothesis is correct, we should observe a higher rate of imitation in older boys than in younger boys, whereas we should not observe such a difference between older and younger girls.

We indeed observed differences in imitation score between younger and older boys on both the transparent apparatus, U = 146, p = .006, r_{rb} = −.423, and the opaque apparatus, U = 162, p = .015, r_{rb} = −.36. We did not observe differences in imitation score between younger and older girls for the transparent apparatus, U = 192.5, p = .296, or for the opaque apparatus, U = 215.5, p = .538. All tests were one-tailed, testing the hypothesis that imitative fidelity is higher in older children than in younger children. Taken together, they indicate that girls probably develop fidelity of imitation necessary for the task by 3 years of age, although deceleration thereafter seems to predominate over continuing development of imitative fidelity. In marked contrast, boys clearly show a continuous developmental trajectory from 3 to 6 years of age, reflecting later enculturation compared with girls.

Table 1
Distribution of successful reward extraction by boys and girls, split by age groups, and test for the difference in proportion of reward extraction between boys and girls.

	Opaque apparatus			Transparent apparatus		
	Success	Failure	Fischer's exact test for difference between sexes frequencies	Success	Failure	Fischer's exact test for difference between sexes frequencies
Younger girls	16	1	Odds ratio = 0.134, p = .106, distribution of proportion success is not different between groups	16	1	Odds ratio = 0.167, p = .113, distribution of proportion success is not different between groups
Younger boys	15	7		16	6	
Older girls	22	3	Odds ratio = 2.864, p = .611, distribution of proportion success is not different between groups	21	4	Odds ratio = 4.000, p = .352, distribution of proportion success is not different between groups
Older boys	21	1		21	1	

Finally, we tested whether the fidelity of imitation could be a predictor of successful retrieval of the reward. Table 1 represents the frequencies of success for younger and older boys and girls for the transparent and opaque apparatuses. Fischer's exact test for odds ratios shows that the proportion of success was not different between boys and girls in either the younger group or the older group, and both were very good at solving the task.

To test the prediction of imitation fidelity on successful extraction of the reward, we ran a logistic regression with age (in months) and imitation score as predictors (sex was not introduced as an additional factor because we did not find significant differences in the proportion of successful reward extraction between boys and girls; this decision was also justified by higher AIC [Akaike information criterion] and BIC [Bayesian information criterion] values for the model fit when sex was added as a factor, with sex reducing the fit of the model to the data both for H1 [the alternative hypothesis] and, if included in the null hypothesis, for H0). For the opaque apparatus, we noted the effect of imitation fidelity on the probability of reward extraction. The model with the predictor was significantly better than the null, $\chi^2(84) = 34.395$, $p < .001$ (McFadden $R^2 = .469$ indicates a very good model fit for H1). The model correctly classified 86% of cases. However, only the imitation fidelity was found to be a significant predictor, $p < .001$, odds ratio = 2.685 (higher imitation fidelity was associated with substantially higher probability of reward extraction). Age (in months), however, was not a significant predictor, $p = .099$, odds ratio = 1.066 (older children were slightly more likely to extract the reward, but the relation between age and proportion of success did not reach the significance level). For the transparent apparatus, we noted an effect of imitation fidelity on the probability of reward extraction. The model with the predictor was significantly better than the null, $\chi^2(84) = 37.398$, $p < .001$ (McFadden $R^2 = .51$ indicates a very good model fit for H1). The model correctly classified 87.4% of cases. However, only the imitation fidelity was found to be a significant predictor, $p < .001$, odds ratio = 2.806 (higher imitation fidelity was associated with substantially higher probability of reward extraction). Age (in months) was not a significant predictor, $p = .309$, odds ratio = 1.039 (older children were slightly more likely to extract the reward, but the relation between age and proportion of success was random). This result suggests that the proportion of success did not significantly increase with age, and both younger and older children were good at reward extraction. Yet, better imitators were somewhat more likely to succeed.

Discussion

The data show that higher imitation fidelity was associated with older age in boys, yet such an association was absent in girls. We provided some statistical evidence suggesting that imitation in girls develops faster than in boys and does not increase significantly from 3 or 4 years to 6 years of age. On the other hand, boys seem to experience a more protracted maturation, which we managed to capture in our study. In addition, we gathered some evidence pointing to imitation fidelity being higher when the apparatus is transparent, and the irrelevancy of some actions performed by the demonstrator can be unequivocally established. We discuss our findings below, beginning with consideration of the different rates of imitation as a function of apparatus type.

As pointed out by Hoehl et al. (2019), over-imitation seems inappropriate with regard to imitation of those opaque actions where utility in resolving tasks cannot be ascertained and is best reserved for cases where experimental participants know for certain that an action they reproduce is ineffectual; our results suggest that this might be an important distinction. In our sample, higher rates of imitation were triggered when the transparency of the apparatus allowed some actions to be scored as irrelevant (e.g., poking in the top hole). A possible explanation for this unexpected finding is that lacking explicit instructions as to whether to copy the style or the goal, children interpreted the model's actions as "ritualistic" (Whitehouse, 2021; Wilks et al., 2016), and hence they focused on fidelity of copying (style) rather than on achievement of the goal (extracting the reward) (Whitehouse, 2021). However, we need to caution the reader that (a) the evidence we have is statistically weak and (b) to avoid the possibility of children discovering the causal underpinnings of the apparatuses right from the beginning, we presented them in a fixed order.

The increase in imitation fidelity by age reported for boys confirms our initial hypothesis and concurs with the theoretical assumptions, replicating empirical findings gathered by other authors (Hoehl et al., 2019; Kenward et al., 2011; Keupp et al., 2013; Marsh et al., 2014; Pittet et al., 2022). Marsh et al. (2014) used familiar objects and reduced to a bare minimum the complexity of the to-be-imitated task to discard causal misattribution as the explanation for over-imitating irrelevant actions (Lyons et al., 2007, 2011). They found that imitation of unnecessary actions increased from 5 to 6 years of age in their sample and that it was precisely those behaviors children rated as irrelevant that prompted more imitation, thereby underlining the social—as opposed to instrumental—character of this over-imitation. Dragon and Poulin-Dubois (2023) also found significant differences between children aged 5 and 3.5 years; the former imitated more unnecessary actions, previously demonstrated by a puppet, while trying to open a plastic box than did the latter. Clay et al. (2018) described a more nuanced relationship between age and over-imitation. They showed 4- to 6-year-old children how to open a box by turning a key in a lock, but, before showing that, they had offered them verbal cues that primed the instrumental or normative aspects of the task and performed arbitrary actions with their elbow (touching the box or placing it above it) or hands (mimicking a sawing motion). Over-imitation of younger children was more sensitive to instrumental priming than to normative cuing and vice versa in the case of the older children. A similar study was conducted by Moraru et al. (2016) in 3- to 6-year-old children where again verbal instructions were used prior or posterior to task demonstration in order to frame the situation to be solved as instrumental or conventional, and they even included the use of direct instructions such as "avoid silly actions"; the task administered here was the puzzle box as used by Horner and Whiten (2005). The main result was that older children tended more to over-imitate regardless of the contextual circumstances, whereas younger participants over-imitated mostly when the conventional aspects of the task had been cued but focused on the causality of the task when they had not. All the aforementioned studies differed in the type of motoric action patterns employed to solve the task, the using or not of verbal cueing, the deployment or not of tools, and so on and yet, notwithstanding all these methodological differences, a common pattern arises: The older the children, the more their imitation was driven by social—as opposed to causal—factors.

On the other hand, the stasis in imitation fidelity across time in girls was unexpected and interesting in equal measure. We hypothesize that girls by 3 years of age might be sufficiently mature in their normative development to feel compelled to imitate the model, which implies that the normative force attributed to model actions by 3-year-old female children is not substantially different from that characterizing 5- or 6-year-olds. Our data lend support to this hypothesis given that no differences were detected between imitation fidelity in 3- and 4-year-old girls as compared with 5- and 6-year-old girls. This is not altogether surprising given that a recent study (Essler et al., 2023) found that imitation of behavior in 18-month-old infants helps them to interiorize social norms, thereby confirming that socialization in relation to norms might start very early. We lack information about the male to female ratios in Essler et al.'s (2023) study to check whether the distress associated with norm violations that the authors targeted was stronger in girls or even privy to this gender. It is possible that changes in imitation fidelity in our study might have emerged in girls if 8-year-olds had been included in it because children at this age are reported to over-imitate more than 5-year-olds (Marsh et al., 2014).

Horner and Whiten's (2005) task and ours are analogous and required the use of stick-like tools and the deployment of comparable action patterns, and yet Horner and Whiten did not report any differences between sexes in the imitation of unnecessary actions. Perhaps this divergence stems from the way that imitation was coded. In our study, we opted for rating the level of the match with the model's behavior on a continuum from 0 to 7 (e.g., imitative accuracy); it is possible that using continuous coding allowed us to detect changes by sex that would have gone unnoticed had we established discrete demarcations (necessary vs. unnecessary actions). Whatever the explanation, it becomes apparent that the sex differences in imitation fidelity we reported here are not attributable to task manipulative demands but rather are attributable to sociocognitive changes linked to developmental maturation.

Another promising candidate for explaining the stasis in imitation fidelity from 3 to 6 years of age that we reported in girls could be the drive to affiliate (or identify) with the demonstrator (Hobson & Hobson, 2008; Nielsen & Blank, 2011; Over & Carpenter, 2012; Tomasello et al., 2005). This is supported by Nielsen's (2006) study where 24-month-old children were selectively imitating only models who acted socially. It is possible that girls' drive toward bonding is superior to that of boys from earlier in infancy and that this is expressed as an increased fidelity of imitation. Alternatively, but not altogether unrelated, a study by Bussey and Perry (1982) of third- and fourth-grade children (average age = 8 years 11 months) found that both sexes accepted same-sex behavior equally, yet boys tended to reject opposite-sex behavior more often than girls. Perhaps developmental changes in the identification and hence acceptance of the demonstrators' behavior occur in boys that are absent in girls, and the difference is reflected in girls' tendency toward imitating women's actions (most mentors in our study were female). Opposite results were reported by Dragon and Poulin-Dubois (2023), who attempted but failed to link over-imitation to social affiliation experimentally in 3.5- to 5-year-old children, which highlights a need for further investigation. Future studies could tackle specifically how matching the biological sex of a demonstrator with that of a participant may impinge on fidelity of imitation across children's developmental age.

We conclude by advancing another possible explanation for the sex dimorphism reported here that relates directly to gender socialization (Grace et al., 2008; Pitcher & Schultz, 1983; Serbin et al., 1994). It is possible that the advantage of younger girls in imitating the behavior of the model is secondary to boys being more often encouraged to think outside the box because they are prompted to solve mechanical problems that make them think in terms of causal connections more often than girls. If that were so, younger boys could have displayed an increased tendency to emulate rather than imitate the model because they felt confident about finding the solution by themselves. In support of this view, teachers in both primary school and high school have been reported to favor the notion that girls achieve academic success through constant work, whereas boys are perceived as less diligent yet "naturally" clever (Klein, 2007; Skelton, 2005). This idea gains support from a study by Bian et al. (2017), where girls as young as 6 years were more reluctant than boys to endorse members of their own gender with the label "super smart." Gender stereotypes about boys being more competent in engineering and computer sciences are extended widely in our society; however, gender-interest stereotypes consisting in the belief that girls are not interested in engineering and computer studies appear as early as 6 years of age (Master et al., 2021) and might dramatically influence the actual choices girls make later on in pursuing scientific careers. Master et al. (2021) recruited more than 2000 participants and demonstrated that girls tended to act according to the gender-interest stereotype by getting less involved in activities that related to engineering or computer sciences. This study hints at a perverse self-fulfilling prophecy; the lack of girls enrolling in engineering or computer careers is not so much in direct relation to their feelings of competence as to having been led to believe from early in life that they lack interest in them. Perhaps future studies on imitation and gender could ascertain just how confident participants feel in finding the solution without relying on a demonstration beforehand.

To summarize, a significant developmental change in fidelity of imitation was detected in male children that was absent in female children, and this effect was statistically very robust. We believe that the most plausible explanation(s) for this dimorphism is (are) (a) an earlier appropriation of social norms taking place in girls that makes them believe that the demonstrator's actions *must* be copied, (b) a stronger identification with the model and increased desire to affiliate with her in younger girls when compared with their male counterparts, and/or (c) younger boys feeling more confident in their ability to find the solution without relying on the model's demonstration. Future follow-up studies are required to gather evidence in support of each of these hypotheses. We think that the divergent developmental courses of imitation that we uncovered here promise to be a source of further insights into the topics of sociocognitive development in girls and boys.

Data availability

Data will be made available on request.

References

Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion*. Rand McNally.

Bem, S. L. (1989). Genital knowledge and gender constancy in preschool children. *Child Development, 60*(3), 649–662. https://doi.org/10.2307/1130730.

Benenson, J. F., Quinn, A., & Stella, S. (2012). Boys affiliate more than girls with a familiar same-sex peer. *Journal of Experimental Child Psychology, 113*(4), 587–593. https://doi.org/10.1016/j.jecp.2012.08.003.

Berl, R. E., & Hewlett, B. S. (2015). Cultural variation in the use of overimitation by the Aka and Ngandu of the Congo Basin. *PLoS One, 10*(3), e120180. https://doi.org/10.1371/journal.pone.0120180.

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science, 355*(6323), 389–391. https://doi.org/10.1126/science.aah6524.

Bussey, K., & Perry, D. G. (1982). Same-sex imitation: The avoidance of cross-sex models or the acceptance of same-sex models? *Sex Roles, 8*, 773–784. https://doi.org/10.1007/BF00287572.

Byrne, R. W., & Russon, A. E. (1998). Learning by imitation: A hierarchical approach. *Behavioral and Brain Sciences, 21*(5), 667–684. https://doi.org/10.1017/S0140525X98001745.

Clay, Z., Over, H., & Tennie, C. (2018). What drives young children to over-imitate? Investigating the effects of age, context, action type, and transitivity. *Journal of Experimental Child Psychology, 166*, 520–534. https://doi.org/10.1016/j.jecp.2017.09.008.

Dragon, M., & Poulin-Dubois, D. (2023). To copy or not to copy: A comparison of selective trust and overimitation in young children. *Cognitive Development, 66*, 101316. https://doi.org/10.1016/j.cogdev.2023.101316.

Duncan, L. E., & Peterson, B. E. (2010). Gender and motivation for achievement, affiliation-intimacy, and power. In J. C. Chrisler & D. R. McCreary (Eds.), *Handbook of gender research in psychology*, Vol. 2: *Gender research in social and applied psychology* (pp. 41–62). Springer Science + Business Media. https://doi.org/10.1007/978-1-4419-1467-5_3.

Essler, S., Christner, N., Becher, T., & Paulus, M. (2023). The ontogenetic emergence of normativity: How action imitation relates to infants' norm enforcement. *Journal of Experimental Child Psychology, 227*, 105591. https://doi.org/10.1016/j.jecp.2022.105591.

Fong, F. T. K., Redshaw, J., & Nielsen, M. (2021). A cross-cultural comparison: Imitation of inefficient tool use in Australian and Malaysian preschool children. *Journal of Cultural Cognitive Science, 5*, 37–47. https://doi.org/10.1007/s41809-021-00077-6.

Frick, A., Clément, F., & Gruber, T. (2017). Evidence for a sex effect during overimitation: Boys copy irrelevant modelled actions more than girls across cultures. *Royal Society Open Science, 4*(12), 170367. https://doi.org/10.1098/rsos.170367.

Gilpin, A. R. (1993). Table for conversion of Kendall's tau to Spearman's rho within the context measures of magnitude of effect for meta-analysis. *Educational and Psychological Measurement, 53*(1), 87–92. https://doi.org/10.1177/0013164493053001007.

Grace, D. M., David, B. J., & Ryan, M. K. (2008). Investigating preschoolers' categorical thinking about gender through imitation, attention, and the use of self-categories. *Child Development, 79*(6), 1928–1941. https://doi.org/10.1111/j.1467-8624.2008.01234.x.

Grueneisen, S., & Tomasello, M. (2017). Children coordinate in a recurrent social dilemma by taking turns and along dominance asymmetries. *Developmental Psychology, 53*(2), 265–273. https://doi.org/10.1037/dev0000236.

Hartup, W. (1983). Peer relations. In P. H. Mussen (Series Ed.) & E. M. Hetherington (Vol. Ed.), *Handbook of child psychology*, Vol. 4: *Socialization, personality and social development* (pp. 103–196). John Wiley.

Heyes, C. (2012). What's social about social learning? *Journal of Comparative Psychology, 126*(2), 193–202. https://doi.org/10.1037/a0025180.

Hobson, R. P., & Hobson, J. A. (2008). Dissociable aspects of imitation: A study in autism. *Journal of Experimental Child Psychology, 101*(3), 170–185. https://doi.org/10.1016/j.jecp.2008.04.007.

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. https://doi.org/10.1016/j.dr.2018.12.002.

Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). *Animal Cognition, 8*(3), 164–181. https://doi.org/10.1007/s10071-004-0239-6.

Kenward, B., Karlsson, M., & Persson, J. (2011). Over-imitation is better explained by norm learning than by distorted causal learning. *Proceedings of the Royal Society B: Biological Sciences, 278*(1709), 1239–1246. https://doi.org/10.1098/rspb.2010.1399.

Keupp, S., Behne, T., & Rakoczy, H. (2013). Why do children overimitate? Normativity is crucial. *Journal of Experimental Child Psychology, 116*(2), 392–406. https://doi.org/10.1016/j.jecp.2013.07.002.

Klein, S. (2007). *Handbook for achieving gender equity through education* (2nd ed.). Lawrence Erlbaum.

Ly, A., Verhagen, A. J., & Wagenmakers, E.-J. (2016). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology, 72*, 19–32. https://doi.org/10.1016/j.jmp.2015.06.004.

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. https://doi.org/10.1098/rstb.2010.0335.

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences of the United States of America, 104*(50), 19751–19756. https://doi.org/10.1073/pnas.0704452104.

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford University Press.

Marsh, L. E., Ropar, D., & Hamilton, A. F. D. C. (2014). The social modulation of imitation fidelity in school-age children. *PLoS One, 9*(1), e86127. https://doi.org/10.1371/journal.pone.0086127.

Master, A., Meltzoff, A. N., & Cheryan, S. (2021). Gender stereotypes about interests start early and cause gender disparities in computer science and engineering. *Proceedings of the National Academy of Sciences of the United States of America, 118*(48), e2100030118. https://doi.org/10.1073/pnas.2100030118.

Moraru, C.-A., Gomez, J.-C., & McGuigan, N. (2016). Developmental changes in the influence of conventional and instrumental cues on over-imitation in 3- to 6-year-old children. *Journal of Experimental Child Psychology, 145*, 34–47. https://doi.org/10.1016/j.jecp.2015.11.017.

Myers, L., & Sirois, M. J. (2006). Spearman correlation coefficients, differences between. In *Encyclopedia of statistical sciences*. https://doi.org/10.1002/0471667196.ess5050.pub2.

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.

Nielsen, M., & Blank, C. (2011). Imitation in young children: When who gets copied is more important than what gets copied. *Developmental Psychology, 47*(4), 1050–1053. https://doi.org/10.1037/a0023866.

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), 182–192. https://doi.org/10.1037/a0024555.

Pitcher, E. G., & Schultz, L. H. (1983). *Boys and girls at play: The development of sex roles*. Bergin & Garvey.

Pittet, I., Kojovic, N., Franchini, M., & Schaer, M. (2022). Trajectories of imitation skills in preschoolers with autism spectrum disorders. *Journal of Neurodevelopmental Disorders, 14*(1), 2. https://doi.org/10.1186/s11689-021-09412-y.

Rakoczy, H., Brosche, N., Warnecken, F., & Tomasello, M. (2009). Young children's understanding of the context-relativity of normative rules in conventional games. *British Journal of Developmental Psychology, 27*(2), 445–456. https://doi.org/10.1348/026151008X337752.

Rakoczy, H., & Schmidt, M. F. (2013). The early ontogeny of social norms. *Child Development Perspectives, 7*(1), 17–21. https://doi.org/10.1111/cdep.12010.

Serbin, L. A., Moller, L. C., Gulko, J., Powlishta, K. K., & Colburne, K. A. (1994). The emergence of gender segregation in toddler playgroups. In C. Leaper (Ed.), *Childhood gender segregation: Causes and consequences* (pp. 7–17). Jossey-Bass.

Skelton, C. (2005). Boys and girls in the elementary school. In C. Kelton, B. Francis, & L. Smulyan (Eds.), *The Sage handbook of gender and education* (pp. 139–151). Sage.

Tomasello, M. (1990). Cultural transmission in the tool use and communicatory signaling of chimpanzees? In S. Parker & K. Gibson (Eds.), *"Language" and intelligence in monkeys and apes: Comparative developmental perspectives* (pp. 274–311). Cambridge University Press. https://doi.org/10.1017/CBO9780511665486.012.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*(5), 675–691. https://doi.org/10.1017/S0140525X05000129.

Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions. *British Journal of Developmental Psychology, 29*(1), 124–130. https://doi.org/10.1348/026151010X532888.

Walker, D. (2003). JMASM9: Converting Kendall's tau for correlational or meta-analytic analyses. *Journal of Modern Applied Statistical Methods, 2*(2), 525–530. https://doi.org/10.22237/jmasm/1067646360.

Wang, Z., Fong, F. T. K., & Meltzoff, A. N. (2021). Enhancing same-gender imitation by highlighting gender norms in Chinese preschool children. *British Journal of Developmental Psychology, 39*(1), 133–152. https://doi.org/10.1111/bjdp.12356.

Whitehouse, H. (2021). Overimitation and the ritual stance. In *The ritual animal: Imitation and cohesion in the evolution of social complexity* (pp. 24–52). Oxford Academic. https://doi.org/10.1093/oso/9780199646364.003.0002.

Whiten, A., Horner, V., & Marshall-Pescini, S. (2005). Selective imitation in child and chimpanzee: A window on the construal of others' actions. In S. Hurley & N. Chater (Eds.). *Perspectives on imitation: From neuroscience to social science* (Vol. 1, pp. 263–283). MIT Press.

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.

Wilks, M., Kapitány, R., & Nielsen, M. (2016). Preschool children's learning proclivities: When the ritual stance trumps the instrumental stance. *British Journal of Developmental Psychology, 34*(3), 402–414. https://doi.org/10.1111/bjdp.12139.