Journal of Experimental Child Psychology 203 (2021) 105036

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Optimizing imitation: Examining cognitive factors leading to imitation, overimitation, and goal emulation in preschoolers

Ruth Speidel [a], Laura Zimmermann [a], Lawrie Green [a], Natalie H. Brito [b], Francys Subiaul [c], Rachel Barr [a,*]

[a] Georgetown University, Washington, DC 20007, USA
[b] New York University, New York, NY 10003, USA
[c] The George Washington University, Washington, DC 20052, USA

A R T I C L E   I N F O

Article history:
Received 2 January 2020
Revised 9 October 2020

Keywords:
Imitation
Overimitation
Efficient learning
Goal emulation
Optimal imitation
Preschoolers
Cognitive constraints

A B S T R A C T

Humans imitate patently irrelevant actions known as *overimitation,* and rather than decreasing with age, overimitation increases with age. Whereas most overimitation research has focused on social factors associated with overimitation, comparatively little is known about the cognitive- and task-specific features that influence overimitation. Specifically, developmental contrasts between imitation and overimitation are confounded by the addition of irrelevant actions to causally necessary actions, increasing sequence length, cognitive load, and processing costs—variables known to be age dependent. We constructed a novel puzzle box task such that a four-step imitation, four-step overimitation, and two-step efficient sequence could be demonstrated using the same apparatus on video. In Experiments 1 and 2, 2.5- to 5-year-olds randomly assigned to imitation and overimitation groups performed significantly more target actions than baseline control groups. Rates of imitation and overimitation increased as a function of age, with older preschoolers outperforming younger preschoolers in both conditions. In Experiment 3, preschoolers were shown a video of an efficient two-step demonstration prior to testing. After they responded, they were shown a four-step overimitation video and were tested on the *same* puzzle box. Children imitated the efficient demonstration, but after watching the overimitation video, they also *overimitated* the irrelevant

* Corresponding author.
E-mail address: rfb5@georgetown.edu (R. Barr).

https://doi.org/10.1016/j.jecp.2020.105036

actions. Once again, older children overimitated more than younger children. Together, results show that preschoolers are faithful, flexible, and persistent overimitators. The fidelity and flexibility of overimitation are constrained not only by social factors but also by basic cognitive processes that vary across age groups. As these constraints diminish, overimitation and flexible (optimal) imitation increases.

Introduction

In contrast to other animals, humans seemingly imitate anything and, sometimes, everything they see. Such blanket copying, including copying behaviors that are causally unnecessary and costly (Horner & Whiten, 2005; Whiten et al., 2016), has been called *overimitation* (Lyons, Young, & Keil, 2007). Since Lyons et al. (2007) coined the term overimitation, countless articles have been published in an effort to characterize overimitation in children and, to a limited extent, in adults (Berl & Hewlett, 2015; Flynn & Smith, 2012; McGuigan, 2012; McGuigan, Gladstone, & Cook, 2012; McGuigan, Makinson, & Whiten, 2011; Nielsen & Tomaselli, 2010; Whiten et al., 2016). However, most studies have focused on understanding the social/cultural conditions that modulate overimitation (for a comprehensive review, see Hoehl et al., 2019). The Hoehl et al. (2019) review laid bare how few studies have explored the role that basic cognitive processes (e.g., attention, memory) play in the *development* of high-fidelity imitation and overimitation. For instance, an important feature of overimitation is that, rather than decreasing with age, overimitation *increases* with age (Berl & Hewlett, 2015; McGuigan et al., 2011; McGuigan, Whiten, Flynn, & Horner, 2007; Subiaul & Schilder, 2014; Whiten et al., 2016). McGuigan et al. (2007) presented children aged 3 to 5 years with a video demonstration. They found that 5-year-olds overimitated, but 3-year-olds were more likely to copy only the causally relevant actions or "emulate." In fact, age-related improvements in overimitation continue into adulthood (e.g., Whiten et al., 2016) and are evident in non-Western cultures (Berl & Hewlett, 2015; Nielsen & Tomaselli, 2010).

What accounts for developmental changes in imitation and overimitation? Overimitation researchers, focused on social and cultural explanations, have largely neglected to explore how basic cognitive processes may similarly constrain overimitation (and social learning more broadly).

Memory is likely to mediate both the fidelity and type of social learning mechanisms employed (Barr & Hayne, 1996, 2000; Bauer & Mandler, 1992; Harnick, 1978; Meltzoff, 1990; Simpson & Riggs, 2011). For example, Simpson and Riggs (2011) used a simplified version of the Horner and Whiten (2005) puzzle box that included an irrelevant action (inserting a tool in the top of the box), followed by a relevant action (inserting the tool in the front opening of the box), and finally using the tool to extract the reward. There were two groups: immediate recall and delayed recall. Results showed that, as predicted, children remember more actions in the immediate recall than in the delayed recall. However, children's forgetting was not random. Following delayed recall, children were more likely to omit the irrelevant action least associated with the goal than they were to omit the causally relevant action most associated with the goal.

Besides long-term and short-term memory, working memory (WM), the ability to temporarily store and manipulate information (Baddeley, 1981, 2012), represents a domain-general cognitive skill that constrains the amount of information children learn and copy in a task (Bauer & Hertsgaard, 1993; Bauer, Hertsgaard, Dropik, & Daly, 1998; Bauer & Mandler, 1992; Harnick, 1978; Kemps, De Rammelaere, & Desmet, 2000; Riggs, Simpson, & Potts, 2011; Simpson et al., 2012). But again, there is relatively little research on how WM similarly constrains the amount and *type* of information children copy when they *fail* to overimitate (Dickerson, Gerhardstein, Zack, & Barr, 2013; Moraru, Gomez, & McGuigan, 2016; Simpson et al., 2012).

For instance, children who fail to overimitate might do so because the task exceeds WM capacities. In this case, WM might predict what is most likely to be copied or ignored, potentially explaining age-related differences. In fact, the few studies that have manipulated task complexity have shown that high-fidelity imitation is inversely related to task complexity, measured in terms of number of items (e.g., Barr et al., 2016; Bauer & Mandler, 1992; Moraru et al., 2016). Few studies have examined how WM constraints affect overimitation. In one study, Subiaul and Schilder (2014) directly addressed this question by exploring how varying cognitive loads (high vs. low WM constraints) affected both the amount and type of information young preschoolers chose to copy and/or ignore. They used a touchscreen-based serial learning task where participants needed to respond to three different simultaneously presented pictures (A, B, and C) in a specific order (A > B > C). Children (3- and 4-year-olds) were tested on two-item sequences (low-WM condition) and three-item sequences (high-WM condition). Each sequence was demonstrated by an adult model who touched each picture in the target order (i.e., serial order response). However, the model always tapped one of the pictures twice (i.e., a multitap response that was causally irrelevant). Because children had been trained on the task prior to testing, they knew the task required that pictures be touched only once. Touching them more than once produced no effect. As predicted, children were more likely to copy both target responses (serial order and multitap) in the low-WM condition than in the high-WM condition.

Only a few studies have directly provided children with a situation where they had an opportunity to learn the most efficient way of solving a task and then were provided with the same task but with irrelevant actions, testing whether children would switch from an efficient response to an inefficient overimitative response (Hoehl, Zettersten, Schleihauf, Grätz, & Pauen 2014; Ronfard, Was, & Harris, 2016; Schleihauf, Graetz, Pauen, & Hoehl, 2018; Wood, Kendal, & Flynn, 2013). Hoehl et al. (2014) used a puzzle box that included relevant and irrelevant actions. Children were exposed to two models: one who demonstrated how to retrieve the reward from the puzzle using causally necessary actions as well as unnecessary actions (inefficient model) followed by one who demonstrated only the causally necessary actions (efficient model). Although all children overimitated following the inefficient demonstration, after witnessing the efficient demonstration, they omitted the irrelevant actions they had just executed in the previous test phase. That is, they did not overimitate following an efficient demonstration. In a subsequent study, Schleihauf et al. (2018) reversed the order of presentation of the demonstrations (i.e., Experiment 2) so that children first saw the efficient response and then the inefficient response. When asked to switch from the efficient actions to the overimitation actions, children did *not* overimitate. These results suggest that children consider the efficiency of responses and prefer efficient responses over inefficient responses in a problem-solving task. Finally, Schleihauf et al. (2020) randomly assigned 5- and 6-year-olds to a video with an adult or robot demonstrator. There were three trials: an inefficient one (Trial 1), then an efficient one (Trial 2), and a final inefficient one (Trial 3). Children imitated the inefficient model and switched to the efficient model but then also switched back to the inefficient model response in Trial 3 to retrieve a marble from a puzzle box demonstrated by both human and robot models. In this design, it was only on the third trial that children again imitated an inefficient model. Similarly, Wood et al. (2013) found that children with prior experience of opening an apparatus efficiently still sometimes copied a puppet's irrelevant actions on the same apparatus. In short, these results broadly support the conclusions of Lyons et al. (2007) by showing that overimitation is difficult to extinguish in instrumental tasks.

The current study

The current study allowed for direct comparison of imitation, overimitation, and goal emulation (or efficiency) using the same puzzle box task. Past research has made direct comparison of imitation with overimitation problematic because overimitation conditions typically require children to encode additional actions, thereby increasing task length and cognitive load, compared with the imitation conditions—factors known to depress overimitation (Subiaul & Schilder, 2014). Yet, seeing or directly experiencing how to solve a problem more efficiently appears to similarly depress overimitation. Yet, in these studies, efficiency is left unmeasured (for exceptions, see Hoehl et al., 2014; Schleihauf et al., 2018; Subiaul, Krajkowski, Price, & Etz, 2015). That is, children are supposed to infer the efficient mechanism, but it is not directly demonstrated or assessed.

The current study sought to address whether children would adopt a more costly and inefficient response after imitating the most optimal responses to solve the puzzle box. As with previous studies (Subiaul, Zimmermann, Renner, Schilder, & Barr, 2016), we used a puzzle box with a clear Plexiglass cover, allowing children to see which actions contributed to the goal and which actions did not. For the current study, we used the puzzle box to control task complexity by using a four-step sequence to test imitation (Experiment 1) and overimitation (Experiment 2) with the same apparatus. Thus, the overimitation demonstration did not contain additional steps; rather, in the imitation version three of the four sequence steps were causally relevant, and in the overimitation version one of the four steps (retrieving the star) was causally relevant. In the imitation configuration of the box, the box was locked, making the additional actions causally relevant. We could then directly compare performance on the same cognitive load. In addition, we included tool use in the imitation and overimitation sequences, but the tool was irrelevant in both conditions because the box could be opened more efficiently by hand than with the tool. There are prior reports of increasing rates of overimitation as a function of age, but here we directly compared age-related changes in imitation and overimitation using the same length sequence for both. To test for age-related differences in the current study, we included 2- to 5-year olds and examined whether younger children (<3.5 years) differed from older children (>3.5 years). This procedure allowed us to test whether memory constraints alter imitation, overimitation, or both relative to baseline performance. Finally, we used video models in order to reduce the processing costs associated with attending to a live model while simultaneously attenuating social/pragmatic demands. Whereas McGuigan et al. (2007) used a video that solely portrayed the experimenter's hands as the experimenter manipulated the task, in the current study the experimenter was revealed from the torso up in between demonstrations and provided salient facial and ostensive cues. In this sense, the intentionality of the demonstrator's actions was clearer, whereas the social salience of the demonstration remained lower than that of a live demonstration. Similarly, in a study with 5- and 6-year-olds, children matched their imitative response to a video demonstration of inefficient and efficient strategies to retrieve a marble from a puzzle box demonstrated by both human and robot models (Schleihauf et al., 2020). These authors argued that social motivations may play a role because children performed causally irrelevant actions when they knew a more efficient response to retrieve the marble. The use of the video model also ensured that demonstrations were consistent across participants.

The experiments were approved by the George Washington and Georgetown University institutional review boards and were operated under the guidelines of the Declaration of Helsinki and its further amendments.

Experiment 1: Imitation

A puzzle box task (reported in Subiaul et al., 2016) was used to assess how toddlers and preschoolers learn after viewing a video demonstration showing a sequence of actions to achieve a goal. Children were randomly assigned to the video model or baseline control group.

Question 1: Will children imitate how to open a novel puzzle box from a video demonstration? We predicted that children would learn from the video relative to baseline but that there would be age-related increases in performance consistent with prior imitation studies (e.g., Dickerson et al., 2013; Subiaul et al., 2015).

Method

Participants

This experiment included 145 typically developing children aged 2 to –5 years (72 female). Children were recruited at the Smithsonian National Zoological Park on the east coast of the United States. These data were previously reported in Subiaul et al. (2016) and were reanalyzed in the current experiment. Most parents (97.9%) reported child ethnicity; of those who reported, 68.1% were White and 31.9% were minority (11.6% Black, 10.9% Asian, 5.1% Hispanic, and 4.3% other). In total, 78 children

(40 female) were randomly assigned to the imitation video demonstration group and 67 children (32 female) were randomly assigned to the baseline control group and received no demonstration. Participants were divided into a younger age group (<42 months; n = 58; M_{age} = 35.29 months, SD = 3.8) and an older age group ($\geq$42 months; n = 87; M_{age} = 49.55 months, SD = 5.1) based on birth dates. An additional 17 children were excluded from the analysis due to failure to complete the task (n = 2), technical error with video (n = 10), experimenter error (n = 1), parental interference (n = 2), or sibling interference (n = 1), ineligible for study due to physical handicap (n = 1). Data were collected from November 2011 to June 2013.

Design

This experiment employed a 2 (Age: older [>42 months] or younger [<42 months]) × 2 (Group: baseline or imitation) between-participants design. Participants were randomly assigned to the imitation or baseline group. Participants in the imitation group were shown a video demonstration of the target actions to find the star in the puzzle box and were then tested with the real puzzle box. Participants in the baseline group did not see the demonstration but were tested with the puzzle box to assess the spontaneous production of the target actions in the absence of a demonstration. There were two boxes: a brown box and a white box. Box assignment was counterbalanced across participants.

Materials

A professional 50-s video of an experimenter demonstrating how to complete each puzzle box task (white or brown box) was developed. In the video, a male adult was shown against a black background, sitting at a table with the target object positioned in front of him. The shots consisted of a wide angle at the beginning (showing the torso of the man at the table with the target object on it while the demonstrator provided ostensive facial and verbal cues), in between each of three identical demonstrations (showing the torso of the man while he provided subsequent noninformative verbal cues), and at the end (showing the torso of the demonstrator smiling, providing a verbal cue, and pointing at the camera). The middle shots were composed of close-ups of the puzzle box on the table and the demonstrator's hands performing the target actions. Cuts were used to transition between the wide-angle and close-up shots, and the visual angle remained straight on for the duration of the video. The verbal cues used were designed to be ostensive yet noninformative: (1) "Look at this!" before Retrieval 1, (2) "Wasn't that fun?" before Retrieval 2, and (3) "One more time!" before Retrieval 3 and "Your turn!" after Retrieval 3. Throughout the demonstration, no parts of the task were labeled and no distractor items were used. Videos for each experiment are publicly available on the Open Science Framework at https://osf.io/gpvdy (Barr, 2020).

Procedure

Participants were recruited and tested in the Think Tank at the Smithsonian National Zoological Park. Parents were asked to complete a general questionnaire (including demographic information, parental education, child language exposure, and media exposure). The entirety of the experiment was video-recorded to allow for subsequent coding of target actions. A summary of the experimental procedures for experiments 1 to 3 is provided in Table 1.

Demonstration phase. Those in the imitation group viewed a 50-s video presented on a 45 × 52.8-cm computer screen of an adult performing the relevant target actions and successfully retrieving the star three times from the brown or white box. If children became distracted during the demonstration, the experimenter redirected their attention to the video. The demonstration phase was followed immediately by a test phase. Children in the baseline condition did not participate in the demonstration phase.

Test phase. The test phase was identical for children in the imitation and baseline groups. Children in the imitation group were given the test stimuli that corresponded to the puzzle box that had been demonstrated in the video. The experimenter placed the three-dimensional (3D) version of the box in front of children and repeated "Your turn!" Children were given 30-s from the time they first touched the box to interact with the box. Following the test phase, there was a manipulation check.

Table 1
Summary of experimental procedures used in Experiments 1 to 3.

Condition	Demonstration phase	Test phase
Experiment 1	Imitation puzzle box configuration	
Imitation	Demonstration of causally relevant target actions that are necessary to complete the sequence and retrieve the star from the puzzle box.	Measure whether children copy the imitation condition target actions demonstrated.
Baseline: Imitation	No demonstration of the target actions.	Measure whether children spontaneously produce target actions.
Experiment 2	Overimitation puzzle box configuration	
Overimitation	Demonstration of target actions, some of which are necessary to retrieve the star from the puzzle box. Other actions are irrelevant.	Measure whether children copy the overimitation of target actions.
Baseline: Overimitation	No demonstration of the target actions.	Measure whether children spontaneously produce target actions.
Experiment 3	Overimitation puzzle box configuration	
Efficiency	Demonstration of necessary target actions on the overimitation puzzle box.	Measure whether children copy the efficiency target actions demonstrated.
Baseline: Efficiency	No demonstration of the target actions.	Measure whether infants spontaneously produce efficiency target actions.

During the manipulation check, the experimenter performed a live demonstration of the same sequence of target actions demonstrated in the video, and children were given another turn to find the star. The manipulation check ensured that children of all ages had the motoric ability to manipulate the puzzle box. All children did. The white box target action sequence is shown in Fig. 1.

Coding

We calculated the following raw score measures of the following action types: target actions, order scores, tool use, distractor errors, and return errors. We also calculated the following composite score measures: goal emulation and fidelity composite. A description of each measure can be found in Table 2.

Note that Steps 1 and 2 in the imitation condition unlock the puzzle box and therefore are necessary, whereas in the overimitation condition the box is never locked and so both steps are not causally relevant. In Step 3, however, the use of the tool to push or pull the box with the stick is included as a target action in both the imitation and overimitation conditions because it is causally relevant but inefficient; consequently, we classified tool use as overimitation in both conditions. Children could pull the box open with their hands, as is demonstrated in the efficient condition. We compared whether children added inefficient tool use in Experiment 3 after the efficient demonstration. Tool use is marked as “N/A” in the efficiency condition, but we coded it if it occurred as a distractor error. Tool use did not occur spontaneously in any baseline condition.

- *Target actions:* These were defined as the total number of demonstrated actions (regardless of causal relevance) reproduced by children regardless of the order in which they were performed. This measure assessed children’s free recall of the four actions associated with retrieving the star from inside the box.
- *Order scores:* In contrast to the target action score, this measure evaluated children’s ability to encode both the actions and their serial order.
- *Tool use.* This measure assessed children’s ability to use the tool in the way it was demonstrated.
- *Distractor errors:* These represented use of a tool or a piece of Velcro that was not demonstrated in the video.
- *Return errors:* These corresponded to the sum of actions of returning items to their starting locations prior to the completion of the goal.

  Composite score measures:

- *Goal emulation:* This was measured as the number of behaviors associated with retrieving the star that were *not* demonstrated in the video (cf. Table 2).
- *Fidelity composite:* This was a measure of children's recall of target responses (causally relevant and irrelevant) in the demonstrated serial order (target actions + order score − return and distractor errors) (cf. Table 1).

All coders were trained on the box protocol using a reliability criterion of kappa > .70. For Experiment 1, a second independent coder scored 27.7% of the videos to determine the reliability of the ratings. For raw score measures of action types, there was an inter-rater reliability kappa of .89.

Results and discussion

Preliminary analysis

We first conducted a series of analyses examining whether the box or gender of the child or the location of testing (Experiment 2 only) entered into any main effects or interactions. Gender was not significant in any of the three experiments and did not enter into any interactions (Experiment 1: lowest $p = .24$; Experiment 2: lowest $p = .11$; Experiment 3: lowest $p = .19$), and neither was box for Experiments 1 and 3 (Experiment 1: lowest $p = .24$; Experiment 3: lowest $p = .12$) or location for Experiment 2 (lowest $p = .41$), and these variables were not considered further. For Experiment 2, box was a significant factor but did not enter into any significant interactions. Box was included in Experiment 2.

Imitation score: Target actions

The number of target actions for imitation were calculated and compared with baseline. A 2 (Age: older or younger) × 2 (Group: imitation or baseline) analysis of variance (ANOVA) on total target actions was conducted. There was a main effect of group, $F(1, 140) = 163.60$, $p < .001$, partial eta-squared (η_p^2) = .54, showing that the imitation group significantly outperformed the baseline group. There was a main effect of age, $F(1, 140) = 5.98$, $p < .02$, $\eta_p^2 = .04$, and a significant age by group interaction, $F(1, 140) = 5.44$, $p < .03$, $\eta_p^2 = .04$, indicating that older children performed better on the imitation task, but groups did not differ at baseline. These results indicate that, as predicted, children can effectively learn from the video demonstration and that older children performed better on the imitation task. The results from Experiment 1 are shown in Fig. 2A as a function of the proportion of target actions completed. We calculated the proportion of target actions completed in order to compare performance across experiments where the number of target actions varied (imitation and overimitation four actions vs. two actions for efficiency). These results suggest that the puzzle box task is an appropriate imitation task for 2- to 5-year-olds, and by design the puzzle box was modified to measure overimitation in Experiment 2.

Imitation fidelity

We also conducted the same 2 (Age: older or younger) × 2 (Group: imitation or baseline) ANOVA on fidelity composite scores, which took into consideration both action and order scores as well as errors. As with the target action proportions, there was a main effect of group, $F(1, 140) = 157.10$, $p < .001$, $\eta_p^2 = .53$, showing that the imitation group significantly outperformed the baseline group. There was a main effect of age, $F(1, 140) = 9.31$, $p < .003$, $\eta_p^2 = .06$, and a significant age by group interaction, $F(1, 140) = 6.26$, $p < .02$, $\eta_p^2 = .04$, indicating that older children performed better on the imitation task, but groups did not differ at baseline. Fig. 2B shows fidelity patterns across experiments.

We also examined whether there were age-related differences in tool use in the experimental groups. We did not include baseline because spontaneous tool use did not occur. A Pearson chi-square demonstrated that tool use was significantly higher in older children than in younger children, $\chi^2(1) = 6.66$, $p < .01$.

Goal emulation

Finally, a 2 (Age: older or younger) × 2 (Group: imitation or baseline) ANOVA on goal emulation was conducted. There were no main effects or interactions (Fs < 1), indicating that goal emulation

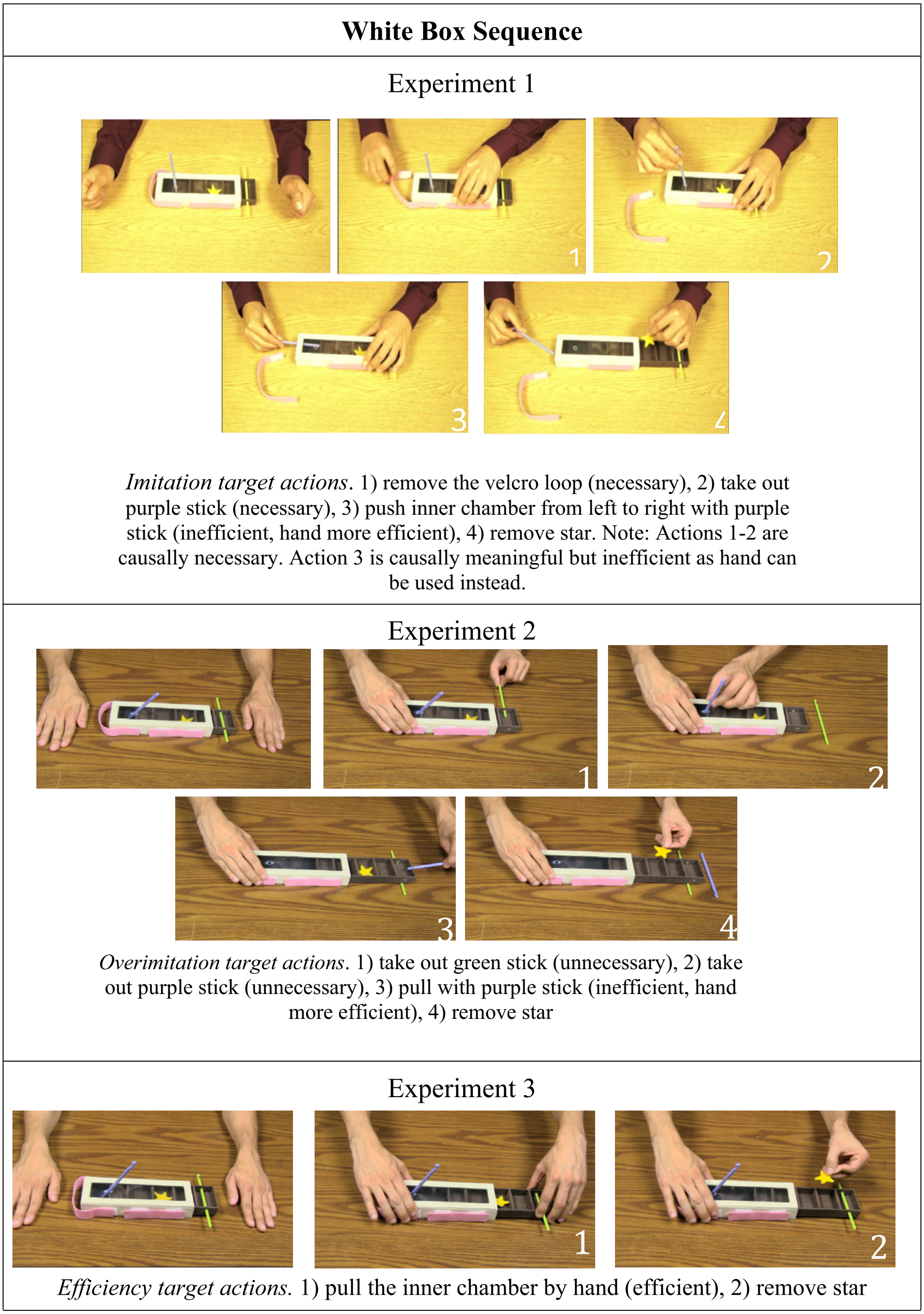

Fig. 1. White box target action sequence for Experiments 1, 2, and 3. (For interpretation of the references to color within this figure, the reader is referred to the Web version of this article.)

Table 2
Puzzle box coding definitions.

Experiment	Imitation (Experiment 1)	Overimitation (Experiments 2 and 3)	Efficiency (Experiment 3)
Target actions (actions explicitly demonstrated in the video regardless of efficiency or causal relevance)	White box: (1) remove the Velcro loop; (2) remove the purple stick; (3) push inner chamber from left to right using the purple stick; (4) remove the yellow felt star. Brown box: (1) remove the green stick; (2) remove the purple stick; (3) push the inner chamber from right to left using the purple stick; (4) remove the red felt star. Max = 4	White box: (1) remove the green stick; (2) remove the purple stick; (3) pull open the box with the purple stick; (4) remove the yellow felt star. Brown box: (1) remove the Velcro loop; (2) remove the purple stick; (3) pull open the box with the purple stick; (4) remove the red felt star. Max = 4	White and brown boxes: (1) pull open box with hand; (2) remove felt star Max = 2
Distractor errors (sum of putting back behaviors, each action = +1)	White box: remove the green stick; remove the distractor side Velcro. Brown box: remove the distractor side or loop Velcro. Max = 2	White box: remove the distractor loop or side Velcro. Brown box: remove the green stick; remove the distractor side Velcro. Max = 2	White and brown boxes: remove Velcro or distractor Velcro; take out purple or green stick Max = 4.
Goal emulation (sum of emulative (nondemonstrated but goal-oriented behaviors, each action = +1)	White and brown boxes: push or pull with hand (2) either purple or green stick in the wrong direction (1) or both sticks at the same time (1); hold up box using gravity (1); pull with purple stick (1); push/pull with purple and/or green stick vertically (2). Max = 10	White and brown boxes: push with purple or green stick vertically (2); push with either stick in the wrong direction (1); pull with both sticks at the same time (1); push or pull with green stick (2); hold up box using gravity (1); push with purple stick (1); pull or push with hand (2). Max = 10	White and brown boxes: hold up box using gravity (1); push with either stick vertically (2) or with either stick in the wrong direction (1); pull with both sticks at the same time (1) or with either stick (2); push with hand (1) or with either stick (2) Max = 10
Return errors (sum of return behaviors)	White and brown boxes: return Velcro; return purple or green stick. Max = 3	White and brown boxes: return Velcro; return purple or green stick. Max = 3	White and brown boxes: return Velcro; return purple or green stick Max = 3
Tool use	White box: Max = 1 (if push with purple or green stick). Brown box: Max = 1 (if push with purple/or green stick). Max = 1	White box: Max = 1 (if pull with purple or green stick). Brown box: Max = 1 (if pull with purple or green stick). Max = 1	Tool use was not demonstrated in the efficiency condition White box: N/A Brown box: N/A Max = N/A
Order score (executing target actions in the demonstrated sequence (even if other actions occur in between)	White/brown box: target actions executed in demonstrated order. Target Action 1 before Target Action 2 (+1); Target Action 2 before Target Action 3 (+1); Target Action 3 before Target Action 4 (+1). Max = 3	White/brown box: target actions executed in demonstrated order. Target Action 1 before Target Action 2 (+1); Target Action 2 before Target Action 3 (+1); Target Action 3 before Target Action 4 (+1). Max = 3	White/brown box: target actions executed in demonstrated order. Max = 1
Fidelity composite	(Target Action Score + Order Score) – (Return Errors + Distractor Errors) Max = 7	(Target Action Score + Order Score) – (Return Errors + Distractor Errors) Max = 7	(Target Action Score + Order Score) – (Return Errors + Distractor Errors) Max = 3

did not differ between baseline and test groups or as a function of age. That is, children did not use an alternative (or idiosyncratic) opening technique to retrieve the reward. Fig. 2C shows goal emulation patterns across experiments.

Experiment 2: Overimitation

Given the successful development of the novel puzzle imitation task in Experiment 1, Experiment 2 adapted the same apparatus to assess overimitation. Identical white and brown boxes were used. However, in Experiment 1, all the demonstrator's actions were causally relevant, if inefficient. In Experiment 2, retrieving the star could be achieved in just two steps using only one's hand (see Fig. 1). Also in Experiment 2, the demonstrator performed a causally irrelevant action in addition to these inefficient actions during the demonstration. Past studies comparing imitation and overimitation have required the encoding and recall of many causally irrelevant and inefficient steps, thereby increasing cognitive load, in order to assess overimitation. This tactic is problematic because it may have resulted in the underestimation of overimitation, especially in younger children who have less developed WM and cognitive resources overall (Best, Miller, & Jones, 2009; Garon, Smith, & Bryson, 2014). The current experiment addressed this concern by using the same number of target actions (four) in both the imitation task and overimitation task so that children were not required to encode and recall additional steps in order to overimitate.

> *Question 2:* Will children overimitate causally irrelevant actions while opening a puzzle box from a video demonstration? Following McGuigan et al.'s (2007) findings, we predicted that children would overimitate causally irrelevant responses from video at rates greater than baseline and that older children would exhibit higher levels of overimitation than younger children.

Method

Participants

A total of 91 typically developing children (42 female, 2 not reported) were recruited at the Smithsonian Institute's National Museum of Natural History Discovery Room and the National Building Museum. Of this sample, 22 children were tested in their homes. Children were randomly assigned to the overimitation experimental or baseline group. The age groups were again divided at 42 months into a group younger than 42 months (n = 46; M_{age} = 35.35 months, SD = 4.37) and a group older than 42 months (n = 45; M_{age} = 51.03 months, SD = 6.06).

Most parents (82.7%) of parents reported child ethnicity; of those who reported, 65.8% were White and 34.2% were minority (19.7% mixed, 3.9% Latino, 3.9% Black, 3.9% Asian, 1.3% Native American, 1.3% other). Parental education was generally high (M = 16.7 years, SD = 1.71, 16% not reported). An additional 14 children were excluded from analysis due to failure to complete the task (n = 10), distraction (n = 1), technical failure (n = 1), or parental interference (n = 2). Data were collected from August 2014 to June 2016.

Design

Independent groups of children were randomly assigned to either the overimitation experimental group or the overimitation control group. It is important to note that a new baseline control group was run for this study using the overimitation puzzle box configuration. This is because in the overimitation box configuration the box is not locked. Half of the children in each group were tested with the white box, and the other half were tested with the brown box (cf. Fig. 1). The boxes in Experiment 2 were the same as those in Experiment 1 except that the locking mechanism was eliminated in Experiment 2, matching actions as closely as possible across experiments. The purple stick and Velcro loop no longer locked the compartment where the star was located, meaning that the first two actions demonstrated were not necessary to retrieve the star. Behavior 3—pulling the box open with the tool—was not necessary in any condition because children could pull the box open with their hand (cf. Fig. 1, Experiment 1). Experimental and baseline control groups were compared.

Apparatus

A 50-s video was produced using identical angles and shots of the experimenter to those used in Experiment 1. The verbal cues and demonstration duration were identical to those used Experiment

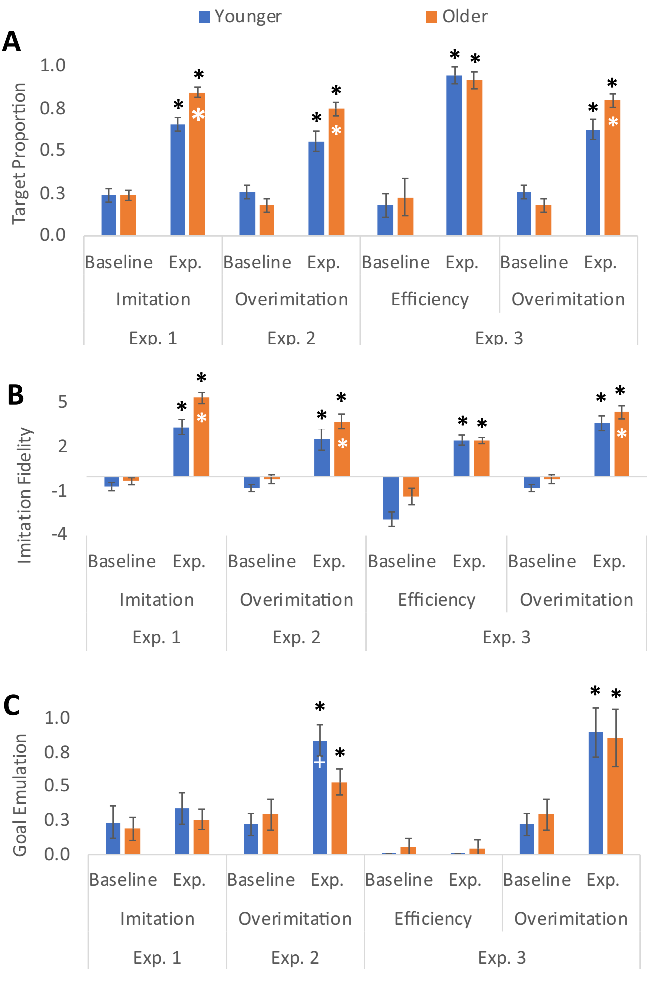

Fig. 2. Results of Experiments 1 to 3: (A) Proportion of target actions across conditions. Younger and older children's average proportions of correct target actions (causally relevant and irrelevant) in the experimental versus baseline groups as a function of experiment are shown. Experiment 1: Younger and older children's performance on the imitation task. Experiment 2: Younger and older children's performance on the overimitation task. Experiment 3: Younger and older children's average efficiency and overimitation target actions in the experimental (efficiency and overimitation) versus baseline groups. (B) Fidelity scores across conditions. Younger and older children's fidelity scores in the experimental versus baseline groups as a function of experiment are shown. (C) Goal emulation across conditions. Younger and older children's goal emulation scores in the experimental versus baseline groups as a function of experiment are shown. Asterisks (*) inside bars represent significant differences between age groups (i.e., younger vs. older), and asterisks above bars represent differences between groups (i.e., baseline vs. experimental). *p < .05; +p < .10.

1. The only aspect of the demonstration that varied was that overimitation actions were modeled rather than imitation actions.

Procedure

Children in the overimitation experimental group viewed a video demonstration of the target actions and were tested in a subsequent 30-s test phase in which the experimenter presented children with the same 3D box and the instruction "It's your turn!" The test phase began when children made first contact with the box. Children in the overimitation baseline control group were given the same box but were not shown a video demonstration of the target actions.

Coding

Coding was identical to that in Experiment 1 except that target actions, goal emulation, and distractor error definitions were altered to reflect the adapted overimitation actions (see Table 2 for details). Two independent coders each coded half of the videos, overlapping on 20% of the videos; for raw score measures of action types, there was an inter-rater reliability kappa of .94.

Results and discussion

Target actions

The number of target actions for the overimitation group was calculated and compared with that for the baseline group. A 2 (Age: older or younger) × 2 (Group: overimitation or overimitation baseline) × 2 (Box: white or brown) ANOVA on total target actions was conducted. There was a main effect of group, $F(1, 83) = 85.16$, $p < .0001$, $\eta_p^2 = .51$, showing that the overimitation group significantly outperformed the overimitation baseline control group. There was no main effect of age, $F(1, 83) = 0.57$, $p = .45$, $\eta_p^2 = .007$, a main effect of box, $F(1, 83) = 17.13$, $p < .001$, $\eta_p^2 = .17$, and a significant age by group interaction, $F(1, 83) = 6.07$, $p = .02$, $\eta_p^2 = .07$. Older children overimitated target actions at a higher rate than younger children in the experimental group. No other interactions were significant (all *p*s > .35). See Fig. 2A.

Imitation fidelity

We also conducted the same 2 (Age: older or younger) × 2 (Group: overimitation or baseline) × 2 (Box: white or brown) ANOVA on fidelity composite scores, which took into consideration both action and order scores as well as errors. As with the target action proportions, there was a main effect of group, $F(1, 83) = 59.98$, $p < .001$, $\eta_p^2 = .42$, showing that the overimitation group significantly outperformed the baseline control group. There was a main effect of age, $F(1, 83) = 5.00$, $p = .03$, $\eta_p^2 = .06$, indicating that older children performed the task with higher fidelity than younger children on the overimitation task, a main effect of box $F(1, 83) = 15.82$, $p < .001$, $\eta_p^2 = .16$, but no age by group interaction, $F(1, 83) = 0.73$, $p = .39$, $\eta_p^2 = .009$. See Fig. 2B.

We also examined whether there were age-related differences in tool use in the experimental groups, again excluding the baseline because spontaneous tool use did not occur. A Pearson chi-square demonstrated that tool use was marginally higher in older children than in younger children, $\chi^2(1) = 3.81$, $p < .051$.

Goal emulation

Finally, a 2 (Age: older or younger) × 2 (Group: overimitation or baseline) ANOVA on goal emulation, defined as the sum of emulative (nondemonstrated but goal-oriented) behaviors (see Table 2 for specific behaviors for each box), was conducted. There was no effect of box on goal emulation, and it was not included in the analysis. In contrast to Experiment 1 where there were no differences in goal emulation as a function of age or group, in Experiment 2 when the mechanism to release the box was not necessary, there was a main effect of group, $F(1, 87) = 17.81$, $p < .001$, $\eta_p^2 = .17$, showing that the overimitation experimental group emulated significantly more than the baseline control group. There was no main effect of age, $F(1, 87) = 1.32$, $p = .25$, $\eta_p^2 = .02$, and a trend for an age by group interaction, $F(1, 87) = 3.43$, $p = .07$, $\eta_p^2 = .04$, indicating that although there were no differences at baseline, there was a trend for younger children to have higher goal emulation scores than older children. This finding is consistent with the finding that older children were more likely to overimitate than younger children. In effect, younger children emulated neither because they were better at understanding the causal features of the task nor because they were less attuned to social variables (e.g., norms, conventions) but simply because they had failed to encode and recall all the demonstrated actions. The result was a higher rate of goal emulation as young children attempted to open the box using idiosyncratic means. Fig. 2C shows the emulation scores.

The results of Experiment 2 provide evidence of overimitation from video at rates greater than baseline (see also McGuigan et al., 2007). Through the use of a video demonstration, the social weight of the model was significantly lowered compared with that in a live demonstration. However, despite the lowered social stimulus, children in the experimental overimitation group learned from the video and overimitated, as evidenced by the significantly higher rate of irrelevant target actions relative to baseline.

Experiment 3: Efficiency and overimitation

To explore children's social learning flexibility, Experiment 3 was designed to assess whether or not children switched from being efficient imitators to costly overimitators on the *same* task. In Experiment 2, although the puzzle box task was designed to be causally transparent, baseline levels on the task were low and suggested that the causal mechanism for efficiently retrieving the star remained unknown to children. Prior research has indicated that overimitation rates are higher when causal mechanisms are ambiguous (McGuigan et al., 2007). Thus, it is possible that children encoded the actions demonstrated in the overimitation video as causally necessary for the successful retrieval of the star. Experiment 3 was designed to test whether children would switch from using an efficient imitative response to using a novel—and costly—overimitative response involving causally irrelevant and inefficient actions.

> *Question 3:* Will children who are given explicit causal information and rewarded for using the most efficient actions switch to a more inefficient and costly overimitative response? Consistent with prior work by Hoehl et al. (2019), we predicted that children would overimitate less following an efficient demonstration on the same task.

Method

Participants

A total of 43 typically developing children (19 female) were recruited at the Smithsonian National Museum of Natural History Discovery Room and the National Building Museum. The age groups were again divided at 42 months into a younger group (<42 months; $n = 19$; $M_{age} = 34.3$ months, $SD = 3.9$) and an older group (≥42 months; $n = 24$; $M_{age} = 47.9$ months, $SD = 6.1$). Most parents (90.1%) reported child ethnicity; of those who reported, 53.8% were White and 46.2% were minority (35.9% mixed, 2.6% Black, and 7.7% Asian). An additional 2 children were excluded from analysis due to failure to complete the task ($n = 1$) or parental interference ($n = 1$). Data were collected from August 2014 to June 2016.

Design

The study employed the same puzzle boxes used in Experiment 2. The overimitation configuration of the puzzle box allowed for efficiency and overimitation demonstrations. The overimitation demonstration was the same as before. The efficiency demonstration involved the quickest and most efficient method to remove the star—pulling the box open using the hand and removing the star. The overimitation control group from Experiment 2 was used for comparison purposes in Experiment 3.

Apparatus

Two 50-s videos were used, employing the same verbal ostensive cues and close-ups as Experiments 1 and 2. One "efficient" video demonstrated efficient actions. There were three demonstrations. For each box, the two target actions were to (1) pull open the box using the hand and (2) remove the felt star. To maintain consistency, the length of the video demonstrations was kept consistent across experiments. In the efficiency video, the target actions were fewer in number and required less time to complete (two vs. four responses); thus, pauses were employed during the close-up portion with hands and box-only of the demonstration so that its length matched that of the imitation and overimitation demonstrations. The second "overimitation" video was identical to that used in Experiment 2. See Fig. 1.

Procedure

First, children viewed the efficient video demonstration and were given a 15-s test phase in which the experimenter presented children with the same 3D box and said, "It's your turn!" The test phase was shortened from 30 to 15-s for two reasons, namely (a) because there were half as many target actions and (b) because children typically needed only 15-s to complete the task. Next, to assess whether or not children would switch from being efficient imitators to costly overimitators, each child viewed the overimitation video demonstration. Children were then given a 30-s test phase to interact with the same puzzle box. Following the overimitation test, a live manipulation check was conducted using the overimitation actions to ensure that all children had the motoric abilities to complete the task. All did. Following Subiaul et al. (2016) criteria for assessing overimitation and considering the possibility for pragmatic miscommunication (Topál, Gergely, Miklósi, Erdohegyi, & Csibra, 2008), this procedure also guaranteed (a) that children knew what were the most efficient (and causally necessary) actions to achieve the goal and (b) that copying *only* the efficient actions was acceptable and would *not* be penalized.

Coding

For the efficiency test coding, target actions (pull box open and remove star), goal emulation, and distractor error definitions were altered to reflect the new efficiency procedure (see Table 2). It is important to note that the alternative goal emulation behaviors for Experiment 3 were very low (see Fig. 2C). Coding for the overimitation test was identical to that used in Experiment 2. The baseline group from Experiment 2 (n = 44) was also coded for efficiency behaviors. Two independent coders each coded half of the videos and overlapped on 23% of the videos; for action type, there was an inter-rater reliability kappa of .96.

Results and discussion

Target actions: Efficiency

The number of target actions for the efficiency experimental group was calculated and compared with a baseline efficiency score using the same puzzle box and control data collected in Experiment 2 (n = 45). Recall that the boxes used in Experiments 2 and 3 were the same and did not include any locking mechanisms. We established a baseline for the efficiency group by calculating the number of times children in the overimitation baseline condition in Experiment 2 performed the respective efficient target actions.

A 2 (Age: older or younger) × 2 (Group: experimental efficiency or efficiency baseline) ANOVA on efficiency scores was conducted for target actions. Recall that the efficiency baseline was calculated from the baseline group in Experiment 2. There was a main effect of group, $F(1, 83) = 102.46$, $p < .0001$, $\eta_p^2 = .55$. There was no main effect of age, $F(1, 83) < 1$, $\eta_p^2 = .001$, and no significant age by group interaction, $F(1, 83) < 1$, $\eta_p^2 = .005$.

Fidelity: Efficiency

A follow-up 2 (Age: older or younger) × 2 (Group: experimental efficiency or efficiency baseline) ANOVA on fidelity scores also yielded a main effect of group, $F(1, 83) = 112.94$, $p < .0001$, $\eta_p^2 = .58$. The main effect of age approached significance, $F(1, 83) = 3.08$, $p = .08$, $\eta_p^2 = .04$, and the same was true for the age by group interaction, $F(1, 83) = 3.26$, $p = .08$, $\eta_p^2 = .04$. Note that the baseline fidelity score for the younger children was lower, which resulted in the trend-level main effect and interaction.

Goal emulation: Efficiency

Finally, we conducted a 2 (Age: older or younger) × 2 (Group: experimental efficiency or efficiency baseline) ANOVA on goal emulation, and there were no main effects or interactions, which is not surprising given that children readily retrieved the star using the demonstrated target actions.

Both older and younger children performed significantly above baseline and were nearly at ceiling on the efficiency test, as indicated by the target proportion score (see Fig. 2A). The baseline efficiency score was low and the fidelity composite score for the efficiency baseline was negative, suggesting that it was not obvious to children that the box could be pulled open using their hand to remove the star. It is important to note that fidelity was lower in the baseline group, indicating that children—and especially younger children—explored the box during baseline and performed distractor actions.

Target actions: Overimitation

Next, to assess switching from efficiency to overimitation, a 2 (Age: older or younger) × 2 (Group: experimental overimitation or overimitation baseline) ANOVA on overimitation scores was conducted on target actions. There was a main effect of group, $F(1, 83) = 121.10$, $p < .001$, $\eta^2_p = .59$, showing again that the overimitation group significantly outperformed the baseline group. There was no main effect of age, $F(1, 83) < 1$, and a significant age by group interaction, $F(1, 84) = 4.865$, $p = .03$, $\eta^2_p = .06$, suggesting that there was no difference in performance at baseline but that overimitation was higher for older children than for younger children.

Fidelity: Overimitation

Our measure of fidelity yielded a similar pattern of results. A 2 (Age: older or younger) × 2 (Group: experimental overimitation or overimitation baseline) ANOVA on the fidelity composite scores produced a main effect of group, $F(1, 83) = 133.20$, $p < .0001$, $\eta^2_p = .62$, showing again that the overimitation group significantly outperformed the baseline group. There was no main effect of age, $F(1, 83) = 2.43$, $p = .12$, $\eta^2_p = .03$, and no significant age by group interaction, $F(1, 83) < 1$. We also examined whether there were age-related differences in tool use in the experimental groups, again excluding the baseline group because spontaneous tool use did not occur. A Pearson chi-square demonstrated that there were no age-related differences in tool use, $\chi^2(1) = 2.06$, $p = .15$.

Emulation: Overimitation

Finally, a 2 (Age: older or younger) × 2 (Group: experimental overimitation or overimitation baseline) ANOVA on goal emulation scores produced a main effect of group, $F(1, 83) = 15.15$, $p < .0001$, $\eta^2_p = .15$, showing greater rates of emulation in the overimitation group than in the baseline group. There was no main effect of age, $F(1, 83) < 1$, and no significant age by group interaction, $F(1, 83) < 1$.

Cross-experiment analyses

In the final set of analyses we compared performance of the experimental groups across the three experiments as a function of age. First, we tested whether there was a difference in the number of target actions between imitation in Experiment 1 and overimitation in Experiments 2 and 3 as a function of age. Specifically, we explored whether younger and older children differed in their tendency to overimitate versus imitate. A 2 (Age) × 3 (Experiment) ANOVA was conducted on the proportion of target actions. There was a main effect of age, $F(1, 162) = 10.86$, $p = .001$, $\eta^2_p = .06$, with older children performing more target actions than younger children. There was no main effect of experiment, $F(1, 162) = 1.23$, $p = .28$, $\eta^2_p = .02$, showing that imitation and overimitation did not vary across experiments. There was no interaction between age and experiment, $F(1, 88) < 1$.

We examined tool use across experiments using a chi-square. The efficiency condition was not included because tool use was not demonstrated. We did not include the baseline conditions because tool use did not occur spontaneously in any baseline condition. When we conducted a chi-square on tool use across the imitation and overimitation conditions, it was significant, $\chi^2(2) = 14.51$, $p < .001$. As can be seen in Table 3, tool use was more likely to occur during the imitation condition than during the overimitation condition.

Table 3
Tool use as a function of experiment.

Condition	Age	Experiment 1: Imitation	Experiment 2: Overimitation	Experiment 3: Efficiency	Experiment 3: Overimitation
Baseline	Younger	0%	0%	0%	0%
	Older	0%	0%	0%	0%
Experimental	Younger	49%	16%	0%	29%
	Older	77%	43%	0%	50%

We also compared fidelity across experiments. There was a main effect of age, $F(1, 162) = 9.46$, $p = .002$, $\eta_p^2 = .06$, with older children copying with higher fidelity than younger children. There was a trend for a main effect of study, $F(1, 162) = 3.02$, $p = .052$, $\eta_p^2 = .04$. A post hoc ($p < .05$) Student–Newman–Kuhls test was not significant. An inspection of the means indicated that fidelity was slightly higher in Experiment 1 than in Experiment 3 and was slightly higher in Experiment 3 than in Experiment 2. There was no interaction between age and experiment, $F(1, 162) < 1$.

Finally, we conducted a 3 (Experiment) $\times$ 2 (Age) ANOVA on goal emulation. There was no main effect of age, $F(1, 162) = 1.82$, $p = .18$, $\eta_p^2 = .01$, and no age by experiment interaction, $F(1, 162) < 1$, $\eta_p^2 = .001$. There was, however, a main effect of experiment, $F(1, 162) = 11.83$, $p < .0001$, $\eta_p^2 = .13$. A post hoc Student–Newman–Kuhls test ($p < .05$) indicated that emulation in Experiment 1 was lower than in Experiments 2 and 3. However, rates of emulation did not differ between Experiments 2 and 3.

Although these results are consistent with a growing body of evidence showing that overimitation increases with age, imitation rates in Experiment 1 did not differ from overimitation rates in Experiments 2 and 3. Overall developmental increases in basic cognitive processes underlying imitative learning as shown in Experiment 1 likely contribute to corresponding age-related increases in overimitation as shown in Experiments 2 and 3. That is, the general cognitive demands of encoding and retrieving a four-step sequence also contribute to the observed differences in overimitation performance. These basic developmental constraints on memory processing have been largely ignored in the overimitation literature and should be considered carefully in future designs. Based on this analysis, overimitation and imitation may represent a distinction without a difference. But again, this is a question that warrants greater consideration and empirical research.

Contrary to our predictions and results from prior studies (Hoehl et al., 2014), observing and copying the most efficient responses does *not* subsequently increase goal emulation rates and depress imitation fidelity or rates of overimitation in the *same* task. These results are surprising. Whereas co-occurring social (affiliative), cultural (normative), and pragmatic (communicative) factors inherent in live demonstration are expected to increase overimitation, video demonstrations *should* uniformly depress these factors, dampening overimitation (Nielsen, Simcock, & Jenkins, 2008). Moreover, overimitation persisted even after being given firsthand knowledge (observing) and experience (recalling) about the most efficient means of solving a problem. This result indicates that overimitation is a robust and stable feature of human social learning—one that, as other studies have shown, is not easily disrupted (Lyons, Damrosch, Lin, Macris, & Keil, 2011; Lyons et al., 2007; Whiten et al., 2016).

General discussion

These results shed new light on the cognitive underpinnings of overimitation and flexible social learning. Experiment 1 showed that, consistent with prior research (Dickerson et al., 2013; Huang, Heyes, & Charman, 2006; Lauricella, Barr, & Calvert, 2016; McGuigan et al., 2007; Zimmermann et al., 2015), children aged 2.5 to 5 years can effectively learn to imitate from a video demonstration performing significantly above baseline. In addition, like various other studies (for a review, see Hoehl et al., 2019), we found age-related differences in imitation performance yet no comparable changes in goal emulation. Experiments 2 and 3 revealed that overimitation does not require a live demonstrator providing ostensive and social cues; children provided with only a video demonstration overimitated much like children provided with a live demonstration. These findings replicate and extend those of Schleihauf et al. (2020), who tested older children. Even with the limited social input intrinsic to a video demonstration, children still displayed the ability to imitate and overimitate novel complex tool-using actions from a video, suggesting that in some cases video might be an effective learning tool. Children also demonstrated the ability to switch strategies after watching a video. Consistent with Carr, Kendal, and Flynn (2015), when social information is delivered via video demonstration, it represents a robust social learning cue. In addition, as with prior studies, there were significant age-related increases in overimitation (for a review, see Hoehl et al., 2019). We also found high levels of emulation relative to baseline across age groups. Children's emulation was not idiosyncratic but rather goal directed. When children emulated, they tended to omit tool use, which corresponded with the most challenging and inefficient action (for similar results, see Nielsen, 2006; Subiaul & Schilder, 2014). Finally, Experiment 3 demonstrated that children were able to flexibly switch from an efficient form of imitation to a less efficient form of imitation, with older children switching more often than younger children.

Despite the large number of overimitation studies with children, surprisingly few have directly explored what accounts for the age-related changes in overimitation almost all of them report. Experiments 1 and 2 demonstrate that when the number of actions match across tasks, there are age-related increases in *both* imitation and overimitation. These findings suggest that there are distinct cognitive constraints on imitation learning in general that may also account for some of the age-related changes in *over*imitation. That is, imitation and overimitation are likely to share common cognitive processes. Overall, when the sequence length (i.e., cognitive load) was equated then imitation and overimitation performance did not differ. However, in overimitation conditions when cognitive load was higher, goal emulation was higher and fidelity was lower. That is, children deployed different social learning processes based on the constraints of the puzzle box task.

However, the fact that overimitation rates did not differ between Experiments 2 and 3 seemingly contradicts this conclusion given that cognitive load was higher in Experiment 3, where children were exposed to competing responses on the same task (efficient and inefficient) relative to Experiment 2, where children were exposed to just one response (inefficient). But recall that the additional responses contributing to the extra load in Experiment 3 were familiar actions performed at ceiling. A result that seemingly contradicts the normative theory of overimitation given that children in the current study received both material rewards (i.e., star) and social rewards (i.e., "Good job!"). These procedures should have depressed overimitation but did not. According to the *normative theory* (Kenward, Karlsson, & Persson, 2011; Keupp, Behne, & Rakoczy, 2013), children should encode the first demonstration (i.e., the efficiency demonstration) as the norm and should reject and refuse to imitate any subsequent demonstrations (i.e., the overimitation demonstration) because they serve as violations of the norm. Other studies that used a comparatively simpler task (involving three familiar responses) showed that varied video demonstrations also resulted in children imitating based on established norms (Evans, Laland, Carpenter, & Kendal, 2018). Although an empirical question, it is possible that children may be more likely to switch responses after observing a video demonstration rather than a live experimenter where the efficient response is more likely to set up the expectation of a norm (Hoehl et al., 2014).

Alternatively, the *automatic causal encoding theory* (Lyons, 2008; Lyons et al., 2007, 2011) suggests that after viewing an actor perform a task intentionally, children will automatically revise and update their causal understanding, encoding all the model's actions as causally meaningful even when previous knowledge indicates otherwise. Our results are more consistent with this theory. Within this framework, there are at least two possible ways to interpret children's performance. One is that on viewing each video demonstration, where a model intentionally manipulates the puzzle box, children revised and updated their causal understanding of the puzzle box, copying all actions whether causally necessary or not. Moreover, in Experiment 3, preschoolers—especially older preschoolers—were willing to incur a cost to retrieve the star from the box via tool use. These results are in contrast to those of Chevalier's (2018) cognitive discounting task. Chevalier found that 7- to 12-year-olds were able to detect a difference in cognitive load both objectively (accuracy and pupil dilation) and subjectively (difficulty rating) on an *n*-back task. Children were significantly more likely to choose the lower cognitive load (e.g., chose to complete the 1-back task instead of the 2-back task) even if it meant earning few points to conserve cognitive effort. Like Chevalier (2018), Hoehl et al. (2014) found that under social learning conditions children were willing to reduce cognitive load, to shift from overimitating to emulating, whereas Schleihauf et al. (2018) found that children were not willing to go in the opposite direction and to shift from emulating to overimitating. In contrast, in the current study, 2- to 5-year-olds were willing to shift from an easier, faster, and more efficient emulative response to retrieve the star to a more effortful, longer overimitative response that involved tool use.

Other factors might contribute to our pattern of results. Marsh, Ropar, and Hamilton (2019) showed that children are more likely to overimitate in challenging novel tasks than in easy familiar tasks. In addition to differences in task complexity, another potential explanation may have to do with differences in the reward structure between the current study and past studies. Prior studies were based on comparative tasks, with children earning a tangible reward each time they solved the puzzle box and incurring a cost if they solved the puzzle box less efficiently (Chevalier, 2018; Hoehl et al., 2014; Schleihauf et al., 2018). For example, in Schleihauf et al. (2018) study, children were allowed to keep the ball that they retrieved from the puzzle box. If they shifted to the slower, riskier overimitation actions, they risked not getting the ball. Finding the ball was an explicit goal that may have led to conditions of optimal emulation. In the current study, the reward was not explicitly tied to finding the star. Children did not keep the star; rather, they collected stickers after the end of the trial regardless of whether they found the star or not. Children appeared to enjoy retrieving the star, as gauged by facial expressions and willingness to retrieve the star on multiple occasions. It is possible that finding the star was an intrinsically rewarding activity for children such that even though the overimitation condition had a high load, children were willing to learn and generate a new response. That is, in the current study, the perceived goal was to learn *how* to retrieve the star. The motivation to learn was the means, and finding the star was an implicit goal leading to conditions for optimal imitation. Children may be more flexible under optimal *imitation* conditions rather than optimal *emulation* conditions. The modulation of imitation and emulation may also be due to more efficient executive functioning (Garon et al., 2014) and age-related differences in the perceived goals of those actions (e.g., Wohlschläger, Gattis, & Bekkering, 2003). That might explain why older children are able to switch or be more flexible than younger children. Other factors such as whether there was one or multiple demonstrators, whether the demonstration was live or on video, and whether gestures (e.g., pedagogical cues) were included or not in the overimitation sections would also need to be controlled. Taken together, these results suggest that object novelty and complexity, perceived goals, and reward structure all may contribute to age-related differences in overimitation, and examination of each of these factors warrants further empirical attention.

After an extensive review of the overimitation literature, Hoehl et al. (2019) similarly concluded that multiple theories explain overimitation, depending on situational demands that include context bias (e.g., type of models) and content bias (e.g., type of task). They concluded that overimitation is typically an adaptive response to situational cues for which it is difficult for children to decipher the causal mechanisms or that increase the likelihood of social or affiliative factors. As evidenced by factors such as the early acquisition of tool use (Casler & Kelemen, 2005), imitation serves as an adaptive avenue for more efficient and complex learning processes. In all three experiments, children in experimental conditions significantly exceeded performance of children in baseline conditions who had not seen demonstrations of the various target actions. In specific case-by-case situations, such as when a demonstrator performs irrelevant steps in the process of achieving an end goal, it may be less efficient to emulate. This is particularly true in ritual contexts that involve causally opaque arbitrary actions that cannot be selectively omitted (Legare, Wen, Herrmann, & Whitehouse, 2015). When tasked to learn from another person, overimitation is still better than the alternative of ignoring all the demonstrator's actions and attempting to complete the task alone—or, in the absence of causal knowledge for each action, deciding which actions to copy and which to ignore. Thus, although overimitation may result in a less efficient learning means for instrumental actions in the short term, it is likely to be an overall effective means of acquiring knowledge. This last point is an important consideration given that most human endeavors involve copying meaningless sequences and actions that are causally opaque (e.g., ritual)—a skill that might represent a human cognitive specialization (Renner, Patterson, & Subiaul, 2020). The current findings contribute to this growing body of literature by highlighting that cognitive constraints such as affordance, task complexity, and reward structure are likely to set the stage for optimal imitation.

Fig. 3 offers a general model of cognitive constraints for understanding and predicting which social learning mechanism children will use under different conditions (i.e., overimitate vs. emulate). This model has three general features believed to constrain imitation: cognitive load, encoding, and retrieval. Cognitive load is anything that increases task complexity and constrains what is encoded and recalled. For example, in the current series of experiments, load was manipulated by the number of items (more items, higher load) and by the affordances of the puzzle box (e.g., affordance of the causal actions increased when the box was locked, decreasing load). In general, the model predicts that high cognitive load conditions limit encoding and retrieval, increasing exploratory goal-directed responses (i.e., "cause and effect" in our model). The result is higher rates of exploration, lower fidelity, and higher emulation rates. Alternatively, in the low cognitive load conditions, encoding and retrieval are high. The

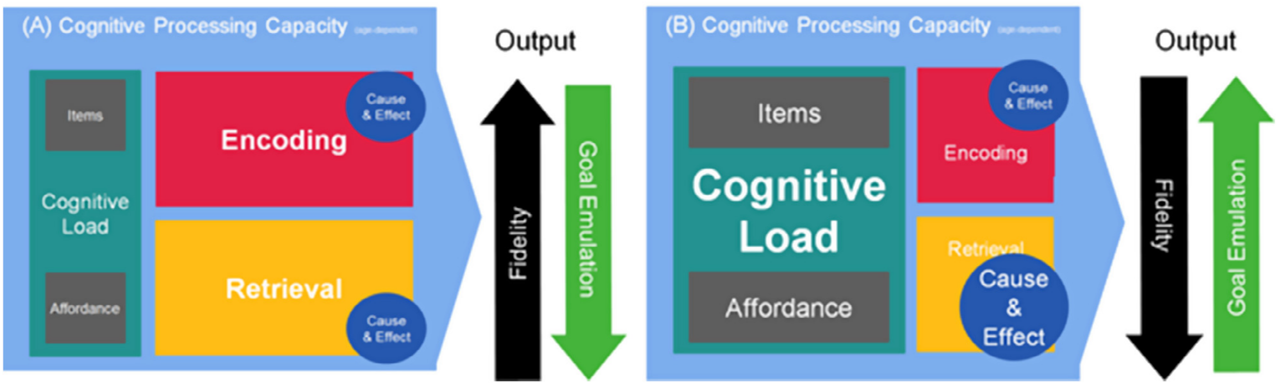

Fig. 3. General cognitive constraint model of optimal imitation. (A) Low cognitive load, more accurate encoding and retrieval, (B) High cognitive load, less accurate encoding and retrieval.

result is more accurate imitation and low emulation rates. Cognitive processing capacity and hence cognitive load constraints are age and experience dependent. So, older individuals will be able to bear larger cognitive loads than younger individuals. In the current study, there were age-related differences in imitation rates when the number of items to be remembered was higher (Experiment 1: imitation; Experiments 2 and 3: overimitation) but not when the number of items to be remembered was low (Experiment 3: efficiency). Goal emulation was lower when the affordances for the causal actions were high (Experiment 1: imitation) and when efficient actions were demonstrated (Experiment 3: efficiency) than when causally irrelevant actions were demonstrated and affordances were lower (Experiments 2 and 3: overimitation). The high rates of goal emulation *and* overimitation in Experiments 2 and 3 suggest that a significant number of children had some understanding of the causal mechanisms of the box in Experiment 2 despite the fact that only in Experiment 3 did children receive explicit information as to which responses were causally relevant. Finally, tool use was more likely to occur in the context of other causally relevant actions (Experiment 1: imitation) than when the other actions were causally irrelevant (Experiments 2 and 3: overimitation). This result suggests that when cognitive load was high, children were less likely to copy tool use, which is an inefficient and motorically difficult response. These constraints were particularly evident in younger children; few copied tool use in overimitation conditions (Experiment 2: 16%; Experiment 3: 29%), yet nearly half (49%) did so in Experiment 1. So, young children's selective tool use was not due to some general motor constraint. However, it is unclear whether children who predominantly evidence goal emulation and those who overimitated had a different causal understanding of the task. It is also unclear whether our proposed cognitive constraints model limits children's causal understanding, responses, or both.

This general model can, of course, be further modulated by social and cultural variables that may alternatively lessen or exacerbate the cognitive load and its resulting effect on encoding and retrieval. So, for example, conventional contexts such as ritual might heighten arousal and, by extension, attention. These changes should *lessen* cognitive load, *increase* cognitive processing capacity, and lead to improvements in encoding and retrieval. The result would be high levels of imitation fidelity and low levels of emulation (Herrmann et al., 2013; Legare et al 2015).

Future studies could use the newly developed puzzle box with its different configurations to further test the flexibility of the emerging social learning system. For example, future research could examine imitation and overimitation within the same individuals testing children on one configuration of each puzzle box. It may provide opportunities to further investigate the relationship between imitation and innovation. Voigt, Pauen, and Bechtel-Kuehne (2019) reported that innovation levels increase when children have enough time to explore and there are multiple materials to use. Subiaul et al. (2015) found that when young children observed demonstrations of different strategies by different experimenters, they were able to combine this information to generate an unobserved (innovative) solution to retrieve multiple hidden objects from a puzzle box. If additional research demonstrates that there are optimal imitation conditions, these findings will have important educational implications to further pedagogical goals to learn one type of response or multiple response types. Optimal imitation conditions may foster intrinsic motivation where the goal is to learn multiple affordances of an object. It may also increase exploratory play and, consequently, creative problem solving or innovation. Under optimal imitation conditions, children may be more willing to incur a cost to reaching the goal in order to learn new information about tools or affordances of an object. As a consequence, the reach of social learning is expanded beyond imitation or overimitation of irrelevant acts to include learning about multiple and novel ways of interacting with objects.

Authors' note

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Baddeley, A. (1981). The concept of working memory: A view of its current state and probable future development. *Cognition, 10*, 17–23.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1–29.
Barr, R., (2020, January 2). Puzzle box task videos. Retrieved from https:/osf.io/gpvdy.
Barr, R., & Hayne, H. (1996). The effect of event structure on imitation in infancy: Practice makes perfect?. *Infant Behavior and Development, 19*, 255–259.
Barr, R., & Hayne, H. (2000). Age-related changes in imitation: Implications for memory development. In C. Rovee-Collier, L. P. Lipsitt, & H. Hayne (Eds.). *Progress in infancy research* (Vol. 1, pp. 21–67). Mahwah, NJ: Lawrence Erlbaum.
Barr, R., Moser, A., Rusnak, S., Zimmermann, L., Dickerson, K., Lee, H., & Gerhardstein, P. (2016). The impact of memory load and perceptual cues on puzzle learning by 2-year-olds. *Developmental Psychobiology, 58*, 817–828.
Bauer, P. J., & Hertsgaard, L. A. (1993). Increasing steps in recall of events: Factors facilitating immediate and long-term memory in 13.5- and 16.5-month-old children. *Child Development, 64*, 1204–1223.
Bauer, P. J., Hertsgaard, L. A., Dropik, P., & Daly, B. P. (1998). When even arbitrary order becomes important: Developments in reliable temporal sequencing of arbitrarily ordered events. *Memory, 6*, 165–198.
Bauer, P., & Mandler, J. M. (1992). Putting the horse before the cart: The use of temporal order in recall of events by one-year-old children. *Developmental Psychology, 28*, 441–452.
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.
Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review, 29*, 180–200.
Carr, K., Kendal, R. L., & Flynn, E. G. (2015). Imitate or innovate? Children's innovation is influenced by the efficacy of observed behavior. *Cognition, 142*, 322–332.
Casler, K., & Kelemen, D. (2005). Young children's rapid learning about artifacts. *Developmental Science, 8*, 472–480.
Chevalier, N. (2018). Willing to think hard? The subjective value of cognitive effort in children. *Child Development, 89*, 1283–1295.
Dickerson, K., Gerhardstein, P., Zack, E., & Barr, R. (2013). Age-related changes in learning across early childhood: A new imitation task. *Developmental Psychobiology, 55*, 719–732.
Evans, C. L., Laland, K. N., Carpenter, M., & Kendal, R. L. (2018). Selective copying of the majority suggests children are broadly "optimal-" rather than "over-" imitators. *Developmental Science, 21* e12637.
Flynn, E., & Smith, K. (2012). Investigating the mechanisms of cultural acquisition. *Social Psychology, 43*, 185–195.
Garon, N., Smith, I. M., & Bryson, S. E. (2014). A novel executive function battery for preschoolers: Sensitivity to age differences. *Child Neuropsychology, 20*, 713–736.
Harnick, F. S. (1978). The relationship between ability level and task difficulty in producing imitation in infants. *Child Development, 49*, 209–212.
Herrmann, P. A., Legare, C. H., Harris, P. L., & Whitehouse, H. (2013). Stick to the script: The effect of witnessing multiple actors on children's imitation. *Cognition, 129*(3), 536–543.
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.
Hoehl, S., Zettersten, M., Schleihauf, H., Grätz, S., & Pauen, P. (2014). The role of social interaction and pedagogical cues for eliciting and reducing overimitation in preschoolers. *Journal of Experimental Child Psychology, 122*, 122–133.
Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Animal Cognition, 8*, 164–181.
Huang, C.-T., Heyes, C., & Charman, T. (2006). Preschoolers' behavioural reenactment of "failed attempts": The roles of intention-reading, emulation and mimicry. *Cognitive Development, 21*, 36–45.
Kemps, E., De Rammelaere, S., & Desmet, T. (2000). The development of working memory: Exploring the complementarity of two models. *Journal of Experimental Child Psychology, 77*, 89–109.
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*, 1239–1246.
Keupp, S., Behne, T., & Rakoczy, H. (2013). Why do children overimitate? Normativity is crucial. *Journal of Experimental Child Psychology, 116*, 392–406.
Lauricella, A. R., Barr, R., & Calvert, S. L. (2016). Toddler learning from video: Effect of matched pedagogical cues. *Infant Behavior and Development, 45*, 22–30.
Legare, C. H., Wen, N. J., Herrmann, P. A., & Whitehouse, H. (2015). Imitative flexibility and the development of cultural learning. *Cognition, 142*, 351–361.
Lyons, D. E. (2008). The rational continuum of human imitation. In J. A. Pineda (Ed.), *Mirror neuron systems* (pp. 77–103). New York, NY: Humana Press.
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*, 1158–1167.
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*, 19751–19756.
Marsh, L. E., Ropar, D., & Hamilton, A. F. C. (2019). Are you watching me? The role of audience and object novelty in overimitation. *Journal of Experimental Child Psychology, 180*, 123–130.
McGuigan, N. (2012). The role of transmission biases in the cultural diffusion of irrelevant actions. *Journal of Comparative Psychology, 126*, 150–160.
McGuigan, N., Gladstone, D., & Cook, L. (2012). Is the cultural transmission of irrelevant tool actions in adult humans (*Homo sapiens*) best explained as the result of an evolved conformist bias? *PLoS One, 7*(12) e50863.
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*, 1–18.
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*, 353–364.
Meltzoff, A. N. (1990). Towards a developmental cognitive science: The implications of cross-modal matching and imitation for the development of representation and memory in infancy. *Annals of the New York Academy of Sciences, 608*, 1–37.

Moraru, C., 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.
Nielsen, M. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology, 42*(3), 555–565.
Nielsen, M., Simcock, G., & Jenkins, L. (2008). The effect of social engagement on 24-month-olds' imitation from live and televised models. *Developmental Science, 11*, 722–731.
Nielsen, M., & Tomaselli, K. (2010). Overimitation in Kalahari Bushman children and the origins of human cultural cognition. *Psychological Science, 21*, 729–736.
Renner, E., Patterson, E., & Subiaul, F. (2020). Specialization in the vicarious learning of novel arbitrary sequences in humans but not orangutans. *Philosophical Transactions of the Royal Society B: Biological Sciences, 375*, 20190442.
Riggs, K. J., Simpson, A., & Potts, T. (2011). The development of visual short-term memory for multifeature items during middle childhood. *Journal of Experimental Child Psychology, 108*, 802–809.
Ronfard, S., Was, A. M., & Harris, P. L. (2016). Children teach methods they could not discover for themselves. *Journal of Experimental Child Psychology, 142*, 107–117.
Schleihauf, H., Graetz, S., Pauen, S., & Hoehl, S. (2018). Contrasting social and cognitive accounts on overimitation: The role of causal transparency and prior experiences. *Child Development, 89*, 1039–1055.
Schleihauf, H., Hoehl, S., Tsvetkova, N., König, A., Mombaur, K., & Pauen, S. (2020). Preschoolers' motivation to over-imitate humans and robots. *Child Development*. https://doi.org/10.1111/cdev.13403.
Simpson, A., & Riggs, K. J. (2011). Three- and 4-year-olds encode modeled actions in two ways leading to immediate imitation and delayed emulation. *Developmental Psychology, 47*, 834–840.
Simpson, A., Riggs, K. J., Beck, S. R., Gorniak, S. L., Wu, Y., Abbott, D., & Diamond, A. (2012). Refining the understanding of inhibitory processes: How response prepotency is created and overcome. *Developmental Science, 15*, 62–73.
Subiaul, F., Krajkowski, E., Price, E. E., & Etz, A. (2015). Imitation by combination: Preschool age children evidence summative imitation in a novel problem-solving task. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01410.
Subiaul, F., & Schilder, B. (2014). Working memory constraints on imitation and emulation. *Journal of Experimental Child Psychology, 128*, 190–200.
Subiaul, F., Zimmermann, L., Renner, E., Schilder, B., & Barr, R. (2016). Defining elemental imitation mechanisms: A comparison of cognitive and motor-spatial imitation learning across object- and computer-based tasks. *Journal of Cognition and Development, 17*, 221–243.
Topál, J., Gergely, G., Miklòsi, A., Erdohegyi, A., & Csibra, G. (2008). Infants' perseverative search errors are induced by pragmatic misinterpretation. *Science, 321*, 1831–1834.
Voigt, B., Pauen, S., & Bechtel-Kuehne, S. (2019). Getting the mouse out of the box: Tool innovation in preschoolers. *Journal of Experimental Child Psychology, 184*, 65–81.
Whiten, A., Allan, G., Devlin, S., Kseib, N., Raw, N., & McGuigan, N. (2016). Social learning in the real-world: "Over-imitation" occurs in both children and adults unaware of participation in an experiment and independently of social interaction. *PLoS One, 11* e159920.
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 of London B: Biological Sciences, 358*, 501–515.
Wood, L. A. N., Kendal, R., & Flynn, E. (2013). Whom do children copy? Model-based biases in social learning. *Developmental Review, 33*, 341–356.
Zimmermann, L., Moser, A., Grenell, A., Dickerson, K., Yao, Q., Gerhardstein, P., & Barr, R. (2015). Do semantic contextual cues facilitate transfer learning from video in toddlers? *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00561.