Cognition 202 (2020) 104320

Contents lists available at ScienceDirect

Cognition

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

Intuitive invention by summative imitation in children and adults

Francys Subiaul[a,*], Margaret A. Stanton[b]

[a] Department of Speech, Language, Hearing Sciences, George Washington University, 2115 G Street, NW, Washington, DC 20001, USA
[b] Department of Psychology, Franklin & Marshall College, P.O. Box 3003, Lancaster, PA 17604, USA

ABSTRACT

Humanity's ability to conquer every corner of the planet rests on our inventiveness. But is this inventiveness best explained by individual problem-solving skills or by our species' exceptional social learning abilities? Using a tower-building task, we show that, on average, 3% of 4–6 year old children ($n = 180$) and adults ($n = 192$) independently combined tower pieces to produce the most optimal tower possible, confirming that preschool age children and adults alike are poor independent inventors. Yet, after observing one or more models generate tower elements separately, both children and adults reproduced the demonstrated elements and spontaneously combined them, producing a novel (unobserved) tower of optimal height, evidence of intuitive invention by summative imitation. These results challenge folk concepts of innovation and corroborate those from mathematical models showing that our species' inventiveness generally arise from social learning rather than individual insights. So, rather than being *sui generis*, human inventions are, broadly, *communis generis*.

1. Introduction

Humans are particularly innovative. But when compared to other animals, human innovations are further distinguished by the fact that they evolve, increasing in complexity and adaptiveness over time (Henrich, 2016; Laland, 2017; Mesoudi, 2011). What explains such cumulative cultural evolution? This question has vexed scientists from fields as different as anthropology (Henrich, Boyd, & Richerson, 2008; Ramsey, Bastian, & van Schaik, 2007), biology (Kolodny, Creanza, & Feldman, 2015; Lewis & Laland, 2012), as well as developmental (Beck, Williams, Cutting, Apperly, & Chappell, 2016; Legare & Nielsen, 2015) and evolutionary psychology (Cosmides & Tooby, 2002; Pinker, 2010). To date, most assume that two broadly independent psychological processes underlie the evolution of human cultural products (Cosmides & Tooby, 2002; Legare & Nielsen, 2015; Pinker, 2010; Ramsey et al., 2007): innovation and imitation. While innovation leads to new behaviors that result in cultural change, imitation is associated with the preservation of existing behaviors through faithful copying. Legare and Nielsen (2015) reason that from an early age these two psychological processes act as "dual engines" leading to uniquely human forms of cultural learning that mediate both the generation and replication of adaptive behaviors across generations.

However, it is clear that our species' potential for imitation and innovation are not equivalent. While humans are exceptional and precocious imitators (Tomasello, 2016), both children (Beck, Apperly, Chappell, Guthrie, & Cutting, 2011; Neldner et al., 2019) and adults (Basalla, 1988; Williams, 2010) are poor independent innovators and problem-solvers after controlling for task difficulty (i.e., requisite cognitive skills) and cultural knowledge (i.e., familiarity with related problems). In fact, the anthropological literature is full of examples of skilled explorers succumbing to the elements after failing to solve basic problems that are readily solved by local populations (Boyd, Richerson, & Henrich, 2011; Henrich, 2016). Yet, humans innovate nonetheless. How?

Computer simulations have addressed these questions by proposing learning mechanisms that effectively combine knowledge, behaviors, or cultural products (Henrich et al., 2008; Lewis & Laland, 2012; Mesoudi, 2015). While some models have emphasized the role of 'happy accidents' or 'lucky leaps' mediated by asocial—individual—learning (Kolodny et al., 2015), virtually all models point to social learning as the dominant driver for both innovations and cumulative cultural evolution. Particularly important in these models are hypothesized mechanisms that lead to the elaboration (Kolodny et al., 2015; Mesoudi, 2015) and/or combination of socially learned responses (Lewis & Laland, 2012; Migliano et al., 2020). Cultural learning studies have corroborated the predictions of these models showing that different forms of social learning can produce cumulative cultural changes (Caldwell & Millen, 2009). Other studies have shown that children and adults combine different semantic facts to produce novel inferences and knowledge (Bauer & Larkina, 2017). However, progress has been limited by the fact that most of these studies have assessed cultural learning using a single response or task demonstrated by one model. All of the studies that have used multiple models (Fay, De Kleine, Walker, & Caldwell, 2019; Herrmann, Legare, Harris, & Whitehouse, 2013; Kempe & Mesoudi, 2014; Muthukrishna, Shulman, Vasilescu, & Henrich, 2014) have presented participants with the *same* target response or variations thereof. In some of these studies participants had the opportunity to, for example,

* Corresponding author.
E-mail address: subiaul@gwu.edu (F. Subiaul).

https://doi.org/10.1016/j.cognition.2020.104320
Received 7 November 2019; Received in revised form 22 April 2020; Accepted 3 May 2020
Available online 04 July 2020

adopt different variations of paper-folding techniques to make a paper airplane (Caldwell & Millen, 2009) or variations of knot-tying techniques (Muthukrishna et al., 2014). The innovation in these studies represent examples of innovation by modification (Mesoudi, 2015) rather than innovation by combination, which involves integrating qualitatively *different* responses or products (Subiaul, Krajkowski, Price, & Etz, 2015). Subiaul et al.'s (2015) task included combining different familiar responses (e.g., lifting or sliding) on a single puzzle box. These responses (e.g., removing velcro, lifting door) were yoked to specific goals (e.g., opening compartment, retrieving sticker). While results showed that young children could copy different familiar responses from multiple models, the task did not afford a continuous measure by which to judge cumulative learning nor opportunities for invention, a sub-type of innovation involving the creation of a novel product as opposed to the modification of an existing object (Ramsey et al., 2007). Building on this body of research, the present study used a novel tower task inspired by the spaghetti tower of Caldwell and colleagues (Caldwell & Millen, 2008) and the tool construction task of Price and colleagues (Price, Lambeth, Schapiro, & Whiten, 2009). Specifically, our tower task (Fig. 1) consisted of discrete pieces (Fig. 1A) that could be productively combined in different ways (Fig. 1B) to build complex structures, only some of which were optimal in terms of structural soundness and height (e.g., Fig. 1C, D). These features made it possible to scale the task's difficulty for children and adults. It also allowed us to assess both quantitative and qualitative differences in children's and adults' inventions (c.f., Table 1). Such direct comparisons are critical to answer questions about the ontogeny of complex cognitive skills—such as innovation—which are assumed to be largely the products of cognitive maturation, formal education or both (Beck et al., 2016; Carr, Kendal, & Flynn, 2016; Heyes, 2018; Neldner et al., 2019; Neldner, Mushin, & Nielsen, 2017).

In Experiment 1, we contrasted an independent invention group—Baseline—that did not receive a demonstration or any social input prior to testing, with two summative learning groups: imitation and emulation. Each group saw two different demonstrations that could be optimally combined to generate a novel product. These summative learning groups varied in the amount of social information provided to participants. Specifically, the summative imitation group saw actions and outcomes, allowing them to copy both. The summative emulation group saw only outcomes, requiring them to independently infer the necessary actions.

We hypothesized the following: First, if summative learning is a mechanism for invention, when compared to Baseline, more participants in the summative learning groups should (a) build tower elements (base, apex)

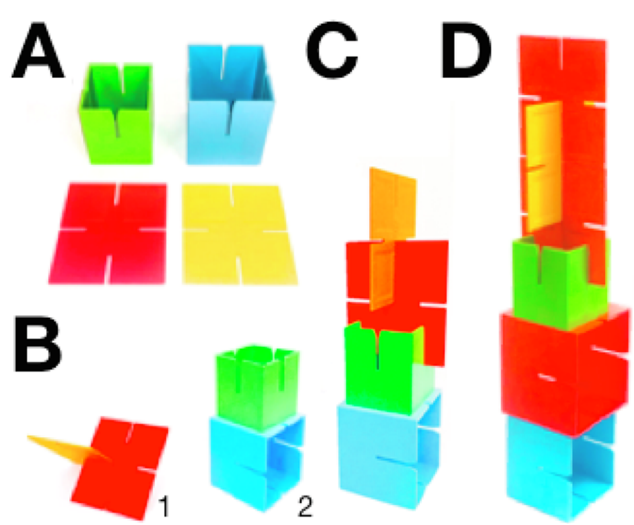

Fig. 1. Tower Task. (A) Child tower pieces: 2 cubes and 2 flat squares. (B) Child tower elements, apex (1), base (2). (C) Child Target (optimal) tower base. (D) Adult tower with 3 flat squares arranged by alternating color and 3 hollow cubes stacked optimally but atypically with middle cube as base. The additional pieces were added to make the task's difficulty comparable across age groups.

and (b) combine them, creating an optimal tower from these component elements (c.f., Table 1). Second, if summative learning contributes to *cumulative* learning, towers produced in the summative groups should become progressively taller across trials and be taller overall than those in Baseline. Third, given the differences in social input between summative imitation (i.e., sees both actions and results) and emulation (i.e., sees only results), we predicted greater rates of intuitive invention in the summative imitation than in the summative emulation group. Finally, if summative learning is an intuitive and unlearned means of inventing, there should be few to no significant differences in the learning patterns of children and adults across learning conditions. All hypotheses, predictions and results can be found in Table 2.

2. Methods: experiment 1

2.1. Participants

Pilot data showed that the frequency of independent invention for the Optimal Tower (i.e., Baseline, see Table 1) among children was < 10%, whereas the frequency of summative learning was < 30%. Given these parameters, the minimal sample size necessary to detect group differences with power = 0.9, *p*-value = .05 and Sampling Ratio = 1.0, was estimated to be 54 (Chow, Shao, Wang, & Lokhnygina, 2017). We based the child and adult sample size on these measures.

2.1.1. Adults

A total of 72 adults (24 per group: Baseline, Imitation, Emulation, Mean_{age} = 22.43 yrs., SD = 6.19, Females = 36) were recruited and tested in the Estelle and Melvin Gelman Library on the Foggy Bottom Campus of The George Washing University in Washington, DC, using GWU IRB approved protocols. For both Exp. 1 and 2, a little over 50% of participants self-described as White/Caucasian (54%), the racial and ethnic breakdown of the remaining participants was: African American = 8%, Asian = 23%, Hispanic = 5%, Mixed = 4%, Did not respond = 6%.

2.1.2. Children

A total of 108 children (36 per group: Baseline, Imitation, Emulation) between the ages of 4–6 years (Mean_{age} = 5.35 yrs., SD = 0.91, Females = 50) were recruited from the National Building Museum in Washington, DC. Children between the ages of 4–6 were selected based on pilot studies showing that children younger than 4 had difficulty combining the flat squares and placing squares in the cubes' ridges. We tested 4- to 6-year olds to evaluate age-related changes in performance among preschool age children prior to significant formal schooling. For both Experiment 1 and 2 approximately 60% of children were White/Caucasian, the racial and ethnic breakdown of the remaining children were as follows: African American = 4%, Asian = 10%, Hispanic = 4%, Mixed = 15%, Did not respond = 8%.

2.2. Task

The Tower Task (Fig. 1) consisted of Dado cubes and Dado squares manufactured by Fat Brain Toy Company based in Elkhorn, NE. The hollow cubes varied in size and color: green cube (6 cm), blue cube (7 cm), red cube (8 cm). The flat squares were the same size (8.8 cm x .25 cm). All the tower pieces had ridges in each side which allowed for the pieces to be connected to each other. The cubes' ridges varied in size: green cube (3 cm x .25 cm), blue cube (3.5 cm x .25 cm), and red cube (4 cm x .25 cm). The ridges in the squares were all the same size: 2.5 cm x.25 cm. Cubes could be stacked to form the base of the tower and squares could be combined to form the apex of the tower.

Because the adult tower included more building pieces, we were able to include causally arbitrary responses that did not affect tower height to explore overimitation or copying of causally arbitrary responses (Horner & Whiten, 2005; Lyons, Young, & Keil, 2007). Specifically, we manipulated the combination of flat squares to form the tower's apex using an

Table 1

Dependent measures and outcomes used in Experiment 1 and 2. Note: Summative learning is a general term that includes two distinct social learning processes that produce inventions, summative imitation (Fig. 2A.1) and summative emulation (Fig. 2A.2).

Measures & outcomes	Optimal (structurally-sound)	Suboptimal (structurally-unstable)
Apex	Linked squares by inserting it in one of the squares' ridges	Linked squares along a solid surface, not in one of the squares' ridges
Base	Rotated cube(s) on side, stacking another cube on solid surface	Balanced one cube atop another's edges
Tower	Combined apex and optimal base	Combined apex and suboptimal base
Summative learning	Combined apex and optimal base following a demonstration	Combined apex and suboptimal base following a demonstration
Independent invention	Combined apex and optimal base in Baseline (without a demonstration)	Combined apex and suboptimal base in Baseline (without a demonstration)

Table 2

Summary of the main hypotheses, predictions, and study outcomes. [A] = Adults, [C] = Children.

Hypotheses	Measured Outcome	Predictions	Outcome of Experiment 1	Outcome of Experiment 2
Summative learning is a mechanism for invention	# of successful participants	a) Tower elements: More in summative learning groups compared to baseline b) Optimal tower: More in summative learning groups compared to baseline c) Optimal tower: More in summative imitation compared to summative emulation	[A] Supported [C] Supported [A] Supported [C] Supported [A] Not Supported [C] Not Supported	[A] Supported [C] Supported [A] Supported [C] Supported [A] Not supported [C] Not supported
Summative learning contributes to cumulative learning	Tower Height	a) Taller across trials b) Taller in summative learning groups compared to baseline group	[A] Supported [C] Partial Support [A] Supported [C] Not Supported	[A] Partial Support [C] Supported [A] Supported [C] Partial Support
Mechanisms underlying summative and non-summative social learning are the same	No. of successful participants	a) Tower elements: No difference between summative and full *imitation* groups b) Optimal tower: No difference between summative and full *imitation* groups c) Tower elements: No difference between summative and full *emulation* groups d) Optimal tower: No difference between summative and full *emulation* groups	— — — —	[A] Supported [C] Supported [A] Supported [C] Supported [A] Partial support [A] Supported
Summative imitation (but not full imitation) serves as an adaptive filter	No. of participats overimitating	a) Patterned linking and atypical stacking: More adults copied both relative to baseline in the full imitation b) More adults copied patterned apex but not atypical base relative to Baseline in summative imitation		[A] Supported [A] Supported
Summative learning is an intuitive means of inventing	No. of successful participants No. and type of errors	a) Tower elements & Optimal tower: No difference between adults and children in baseline and summative imitation conditions b) No difference between adults and children across learning conditions	Supported Partial Support	

idiosyncratic combination of pieces (Red-Yellow-Red) as well as an unorthodox means of stacking cubes to form the tower's base that involved stacking the two other cubes atop the mid-sized (rather than the largest) cube (Fig. 1D). The unorthodox stacking of cubes introduced not just an irrelevant but a potentially maladaptive structural feature. This feature of the task allowed us to then evaluate whether invention via summative learning would attenuate overimitation and serve as a type of 'adaptive filtering' (Enquist & Ghirlanda, 2007), where suboptimal or maladaptive responses are removed or corrected.

2.3. Procedures

Children observed three live demonstrations. Adults watched a single video demonstration (c.f., Supplementary Movies 1.2–4.1). These demonstration procedures were used to maximize learning in both populations. First, while modern adults are regularly exposed to videos and are accustomed to learning from them, preschool age children have less experience learning from videos and are more accustomed to learning from live models. Second, although video provides greater stimulus control than live demonstrations, children suffer from a significant video learning deficit (Choi, Kirkorian, & Pempek, 2018; Dickerson, Gerhardstein, Zack, & Barr, 2013; Moser et al., 2015). The differences in demonstration number have to do with differences in sustained attention between preschool age children and adults. While watching 1 versus 3 demonstrations is unlikely to affect the performance of adults, a single demonstration would have depressed the performance of children (Barr, Muentener, & Garcia, 2007; Barr, Muentener, Garcia, Fujimoto, & Chavez, 2007). In short, these procedural differences made results more comparable across age groups.

Each demonstration was approximately 30 s in length. There were two social groups that assessed summative learning:

- ***Summative Imitation***: This group was provided with the most social input and required the least individual—inferential—learning. The summative imitation group saw two different models: One who built the base of the tower by rotating and stacking cubes atop each other. The other built the apex of the tower by conjoining the two flat squares (Figs. 2-1A). Participants never saw the two tower elements—base and apex—combined. Following each demonstration participants saw each model disassemble the pieces returning them to the starting state. See Suplementary Material and Movie 1.2.

- **Summative Emulation:** The summative emulation group saw the same demonstration as the Summative Imitation group except that

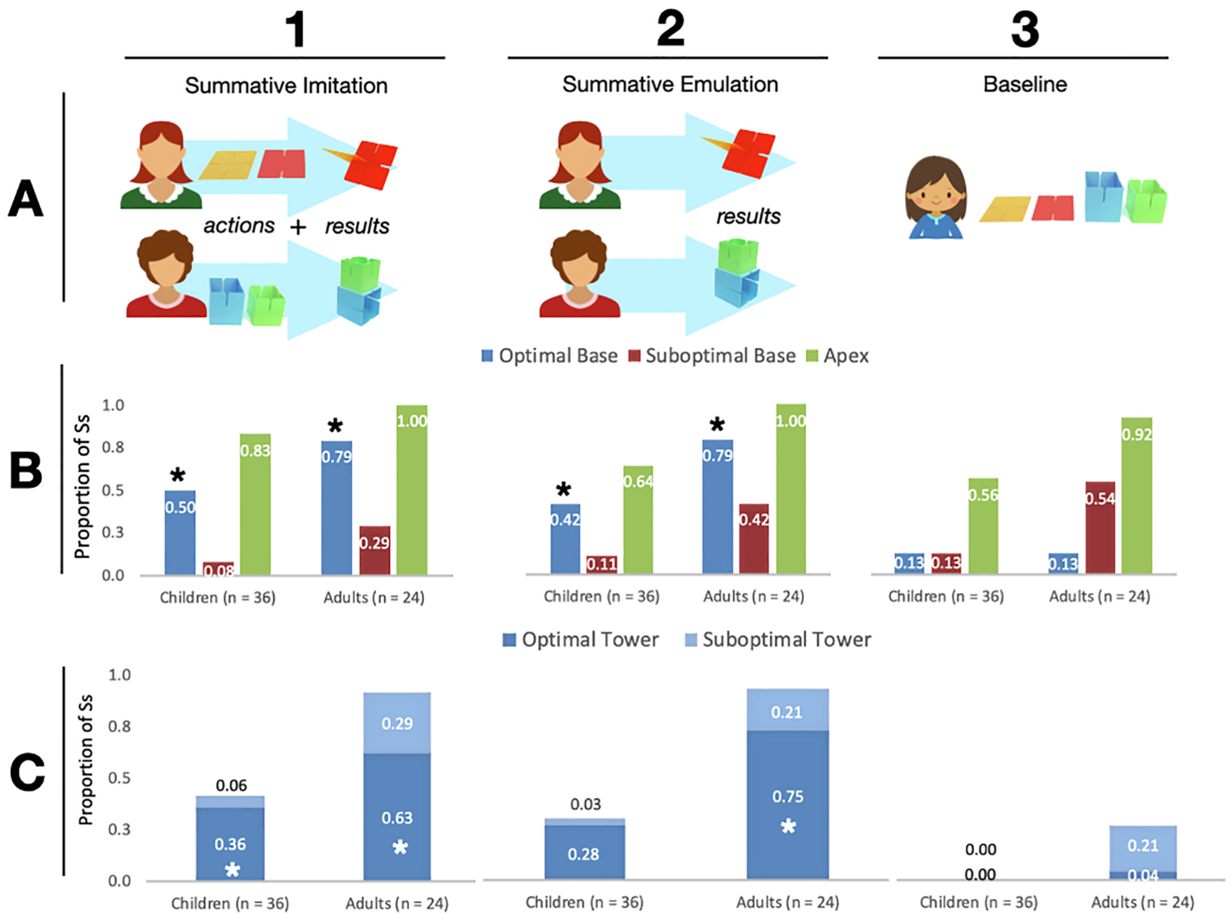

Fig. 2. Experiment 1: Procedures and Results. Rows correspond to (A) experimental groups, (B) proportion of subjects generating tower elements and (C) proportion of subjects generating optimal and suboptimal towers. Columns correspond to each experimental group each measuring a type of summative learning (c.f., Tables 1, 1) two model summative imitation, (2) two model summative emulation and (3) baseline or independent invention. Note: * significantly greater than baseline.

only the assembled tower elements were shown (Fig. 2-A2). The actions used by the models to make each tower element were occluded. Instead of observing the actions, participants only saw completed tower elements—Base/Apex—associated with a specific model (c.f., Suplementary Material and Movie 2.2). In contrast to the summative imitation group, this group was provided with some social input and required to infer the actions necessary to generate the optimal base and apex.

After the demonstration, participants were given the unassembled tower pieces and instructed to build the tallest possible tower with all the pieces.

Social groups were contrasted with the following independent invention group:

- ***Baseline***: This group was provided with no social input and required the most individual—inferential—learning. Testing procedures followed the same procedures used in the summative learning groups except that participants in this group did not observe a demonstration prior to testing.

For adults, the gender of individual participants and models in video matched. However, children always watched female models. We matched the gender of models and participants in adults because research has shown that learning and memory is better when models are matched by sex (Signorella, Bigler, & Liben, 1997). However, we used only female models with children because young children have stranger anxiety, specifically, for unfamiliar males (Heerwagen & Orians, 2002).

During the demonstration, all participants observed the models create specific tower elements. The position of the models (left/right) and associated tower elements were counterbalanced. Demonstrations across all participants and groups used a standard script. Following the demonstration, participants were given 4 (children) or 6 (adult) tower pieces and instructed to build the tallest possible tower using all the pieces. Participants were given an unlimited amount of time to complete a tower. For a trial to end, towers had to stand on their own and all pieces had to be connected or contact each other. When participants finished building the tower, an experimenter measured the tower's height. If participants did not generate a target tower (Child: 22–24 cm; Adult: 34–36 cm)–either optimal, with optimal base (Fig. 1C, D) or suboptimal with suboptimal base (Fig. 3A)–they were given a second trial following the same procedures described above, but no additional demonstration. Before the start of the second trial, the previously built tower was disassembled and tower pieces were laid in front of the participant as in Trial 1. Participants were then told to try building "an even taller tower" (Adult) or "another super tall tower" (Child). If the participant did not generate a target tower, this procedure was repeated on a third and final trial. We used the same protocol for participants in the Baseline group.

2.4. Measures and video coding procedures

We coded all responses made by participants. Because children and adults' towers included a different number of pieces, two coding templates were used, one for each population. The adult template included 47 different responses. The child's template included 31 different

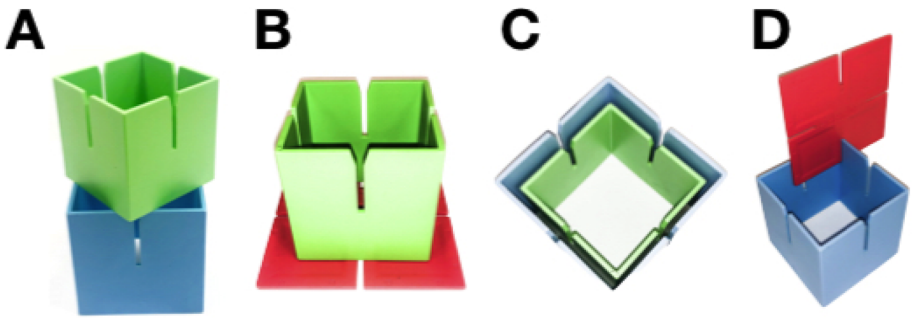

Fig. 3. Suboptimal Responses. (A) suboptimal base, (B) stacking error, (c) Nesting error, (D) balance error.

responses. Target responses were associated with building the optimal tower and included combining all the squares (i.e., Tower's Apex, Fig. 1B-1), rotating cubes on their side and optimally stacking them on a solid, stable surface (i.e. Tower's Base, Fig. 1B-2), as well as combining the two tower elements optimally by inserting squares inside the small cube—(Child, Fig. 1C)—or in either the small or middle cube (Adult, Fig. 1D), or combining the two tower elements sub-optimally (e.g., placing optimally stacked cubes atop combined squares.[1] Table 1 describes the dependent measures used in the present study.

2.4.1. Errors

In addition to coding for these target responses, we also identified three types of errors (Fig. 3A-D). These errors were suboptimal responses that either did not meaningfully contribute to tower height (Nesting and Stacking Errors) or produced very unstable structures (Balancing Errors). The suboptimal base was categorized as a balancing error because it involved balancing cubes on their edges (Fig. 3A), making it structurally unsound and prone to collapse.

Across studies and groups, when participants combined tower elements (apex + base)– optimally or suboptimally–to produce the tallest possible tower, the study ended. The reason for this decision rule being that having discovered the tallest possible tower (despite variation in structural optimality) any other structure would be, necessarily, either the same height or smaller. The only exception was when participants specifically asked for additional opportunities to try to generate an even taller target tower. However, testing sessions never exceeded a total of 3 trials. When participants generated one of the target towers before Trial 3, the responses made on that final trial were reproduced for the remaining incomplete trials. We did this because: (1) leaving cells blank would diminish statistical power when conducting repeated measures and (2) asking subjects to repeat what they did in previous trials would have introduced additional memory and attentional confounds.

Statistical Analyses: All chi-square goodness of fit tests were performed using SPSS 25 (IBM Corp). Bonferroni procedures were used to correct for multiple chi-square tests. All *p*-values were two-tailed. Linear mixed models (LMMs) and generalized linear models (GLMs) were conducted in R version 3.5.2 (Team, 2018) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). Pairwise post hoc comparisons using Tukey's correction for multiple testing were conducted using the emmeans (Lenth, Singmann, Love, Buerkner, & Herve, 2019) package. Significance of fixed effect predictors was determined using likelihood ratio tests (LMMs) and Wald Z tests (GLMs). Model assumptions were evaluated using diagnostic residual plots. Preliminary analyses did not reveal any effects for sex, so we excluded this variable from our analyses.

[1] This type of tower was very rare. However, if it consisted of the optimal base, it was coded as an optimal tower because it was structurally sound. If it consisted of the suboptimal base it was coded as a suboptimal tower. This decision was motivated by the fact that a key outcome measure was summative learning (c.f., Table 1).

2.5. Open practices & data accessibility

Experiment 1 was not preregistered. However, de-identified data for this experiment along with the coding templates used and data analysis scripts are posted here: [https://osf.io/casw8/]. Additional materials used in this study will be made available upon request.

2.6. Results

2.6.1. Was there evidence of summative learning?

To answer this question, we used chi-square goodness of fit tests to compare the number of individuals in each group who produced (i) the optimal base, (ii) the suboptimal base, (iii) the apex, (iv) combined the optimal base and apex to produce the optimal tower and (v) combined the suboptimal base and apex to produce a suboptimal tower. Full Chi-square test results can be found in Supplementary Material 1: Table A1.

Adults' performance differed by group for the optimal base (χ^2 (2) = 29, p < .01, φ = 0.63) with fewer Baseline participants generating the optimal base than either social group (all χ^2 (2) > 8.65, all *ps* < 0.01, all *φs* > 0.35). No other contrast for adults was significant. Results are summarized in Fig. 2-B.

The number of adults generating the optimal tower differed across groups (χ^2 (4) = 18.38, p < .01, φ = 0.5) with more participants in the social groups generating the target tower relative to Baseline (all χ^2 (2) > 16, all *ps* < 0.01, all *φs* > 0.46). No other contrast for adults was significant. Results are summarized in Fig. 2-C.

Like adults, the number of children who built the optimal base differed by group (χ^2 (2) = 12.72, p < .01, all *φs* = 0.34). More children in the social groups produced the optimal tower base, when compared to Baseline (all χ^2 (1) > 8.0, all *ps* < 0.01, all *φs* > 0.27). As can be seen in Fig. 2-B, children and adults alike were more likely to produce the suboptimal tower base in Baseline than in any other group. However, this difference did not reach statistical significance after corrections (c.f., Supplementary Material 1: Table A1).

The number of children who generated the optimal tower also differed by group (χ^2 (2) = 15.75, p < .01, φ = 0.38). Children, like adults, were more likely to produce the optimal tower in the social groups than in Baseline (all χ^2 (1) > 12, all *ps* < 0.01, all *φs* > 0.30). No other contrast for children was significant. Results are summarized in Fig. 2-C.

2.6.2. Does summative learning produce cumulative learning?

To answer this question, we examined predictors of tower height using linear mixed models (LMMs). We created a LMM for adults that included Tower Height as the response variable with trial number (1–3), group (baseline, summative imitation, summative emulation) and the trial by group interaction as fixed effect predictors. Participant ID was included as a random intercept to control for multiple responses from the same participant. Results showed a main effect for group (LRT: χ^2 (2) = 42.97, p < .01) and trial (LRT: χ^2 (2) = 59.71, p < .01). The interaction between trial and group was not significant (LRT: χ^2

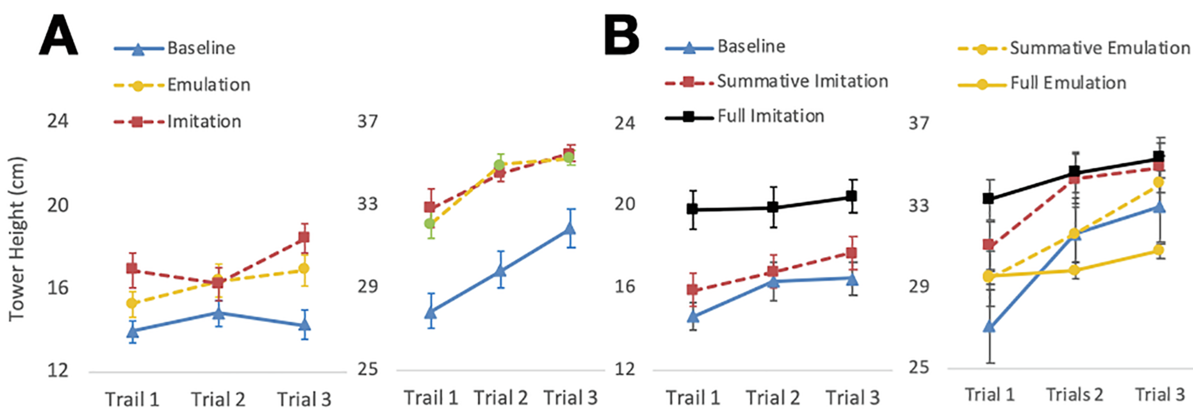

Fig. 4. Cumulative Learning Results: (A) Experiment 1, (B) Experiment 2.

(4) = 3.84, p = .43). Pairwise comparisons showed that towers in Trial 1 were shorter than towers in Trial 2 (p < .01) and Trial 3 (p < .01). Likewise, towers in trial 3 were taller than those of trial 2 (p = .04). Participants in the Baseline group produced significantly shorter towers than those in either the summative imitation (p < .01) or summative emulation (p < .01) groups. The difference between summative learning groups was not significant (p = .97). Results are summarized in Fig. 4A, Supplementary Material 1: Table A5.

We used the same model for children with the exception that the child model included age group (year 4–6). Results showed a main effect for trial (LRT: χ^2 (2) = 11.42, p < .01) and age (LRT: χ^2 (2) = 14.03, p < .01; Supplementary Material 1: Table 3). Tukey's pairwise comparisons showed that towers in trial 3 were taller than towers in Trial 1 (p < .01) and Trial 2 (p = .02). No other comparison was significant; However, group (LRT: χ^2 (2) = 5.41, p = .07) and the trial by group interaction approached statistical significance (LRT: χ^2 (4) = 8.64, p = .07), where 5- and 6-year-olds produced taller towers than 4 - year-olds (p < 01). Like adults, towers in the social groups were taller than those in Baseline. And, while towers in the social groups tended to get taller across trials, those in baseline did not. Results are summarized in Fig. 4A, Supplementary Material 1: Table A4.

2.7. Discussion

Experiment 1 produced several significant findings: First, both adults and young children faithfully replicated two distinct responses on novel objects demonstrated by two different models. Second, participants of all ages spontaneously combined these responses in an unobserved way to generate an optimal tower. Third, there was evidence of cumulative learning across trials as evidenced by significant increases in tower height for both children and adults. However, adults' cumulative learning was more robust relative to children (c.f., Fig. 4). Finally, 11 out of 11 chi-square contrasts between learning groups produced the same results in adults and children (c.f., Supplementary Material 1: Table A1). This pattern of performance suggests that the mechanisms mediating responses in this task are developmentally conserved.

These results, however, raise a number of questions about the underlying mechanisms mediating summative learning in the Tower Task. In an earlier study, Subiaul et al. (2015) found some evidence that imitation fidelity was higher in the two model summative imitation group than in the single model imitation groups. So, Experiment 2 sought to evaluate the following: First, is learning different responses from different models the same as learning different responses from the *same* model? Second, is the combination of different imitated responses (i.e., summative imitation) like the imitation of a single continuous response (i.e., full imitation)? Likewise, is the combination of different emulated responses (i.e., summative emulation) like the emulation of a single continuous response (i.e., full emulation)?

To address these questions, Experiment 2 replicated the procedures used in Experiment 1 with the following differences: First, Experiment 2 compared summative learning with full demonstration groups that do not involve combining different observed responses. And, second, a single model built both the apex and base of the tower. If the cognitive mechanisms underlying summative learning are the same as those that mediate all forms of social learning, there should be no significant differences in performance between the summative and the full demonstration groups. Finally, Experiment 2 explored whether invention by summative imitation and emulation is, itself, a culturally learned, late-developing skill. Cross-cultural studies have found few differences in children's imitation fidelity (Nielsen & Tomaselli, 2010) or overall rates of innovation (Neldner et al., 2017; but see Neldner et al., 2019 for task-specific exceptions). However, prior studies suggest significant developmental discontinuities between the innovation rate of children and adults (Beck et al., 2011). If correct, we might expect both quantitative as well as qualitative differences in children's and adults' patterns of performance across learning groups.

3. Methods: experiment 2

3.1. Participants

3.1.1. Adults

A total of 120 adults (24 per group: Baseline, Summative Imitation, Summative Emulation, Full Imitation, Full Emulation, $Mean_{age}$ = 23.09 yrs., SD = 6.39, Females = 57) were recruited and tested in the Estelle and Melvin Gelman Library on the Foggy Bottom Campus of The George Washington University in Washington, DC using GWU IRB approved protocols.

3.1.2. Children

A total of 72 children (24 per group: Baseline, Summative Imitation, Full Imitation, Mean age = 5.36 yrs., SD = 0.79, Female = 37) were recruited from the National Building Museum in Washington, DC.

3.2. Task: same as in experiment 1

3.2.1. Procedures

Children were randomly assigned to one of three learning groups, either one of two summative learning groups (Table 1)–Summative Imitation (Fig. 5-A1) or Full Imitation (Fig. 5-A2)–or an independent invention (Baseline) group (Table 1, Fig. 5-A3). The procedures used were identical to those used in Experiment 1 with the following exception: Both summative and full imitation groups observed a single model generating each tower element—Base and Apex—and either combining them to show the optimal tower (Full imitation) or not

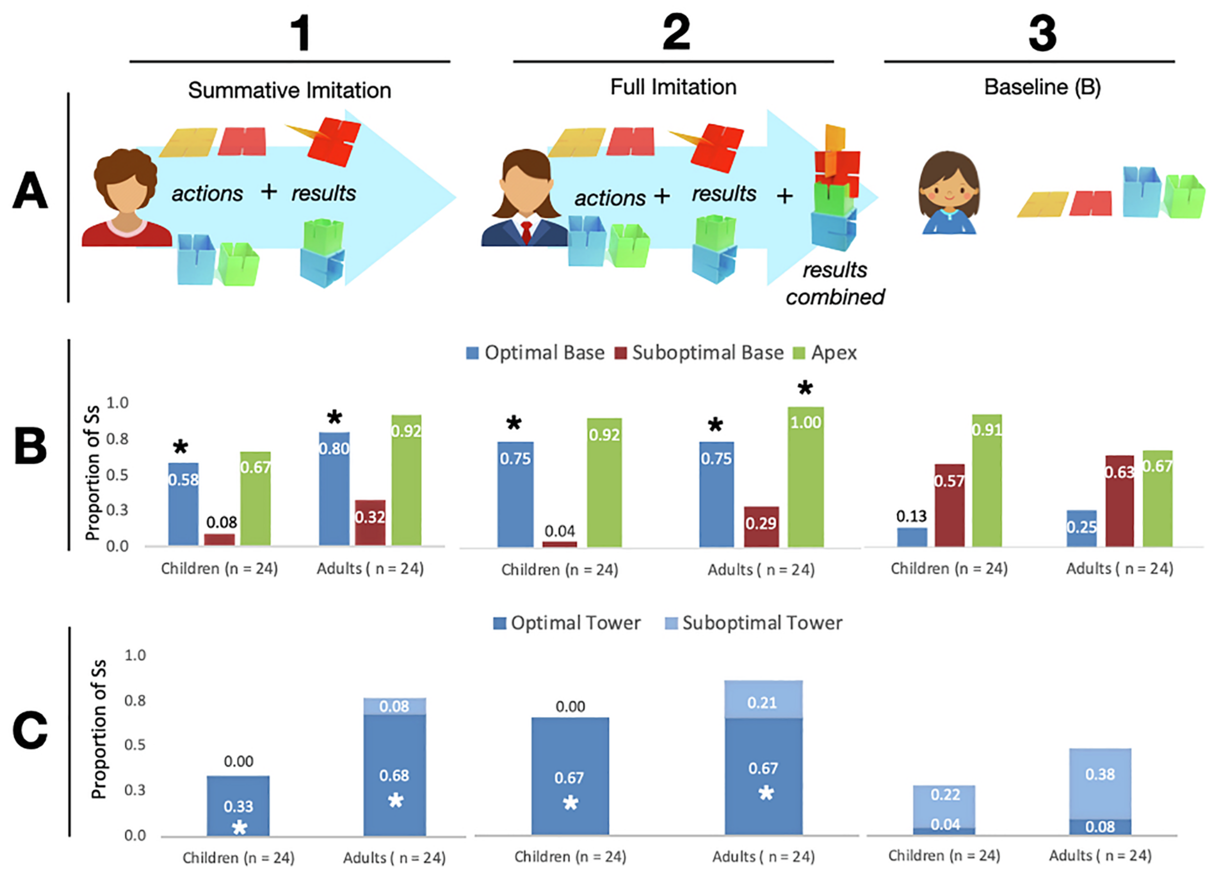

Fig. 5. Experiment 2 Groups and Results for Children and Adults. Rows correspond to (A) experimental groups, (B) proportion of subjects generating tower elements and (C) proportion of subjects generating optimal and suboptimal towers. Columns correspond to each experimental group: (1) one model summative imitation, (2) one model full imitation and (3) baseline. Note: * significantly greater than baseline.

(Summative Imitation).

Adults were randomly assigned to the same three groups as children as well as two other groups: summative emulation (Fig. 6-A1) and full emulation (Fig. 6-A2) group. These groups were identical to the imitation groups with the exception that the actions necessary to produce tower elements (i.e., summative group) or to produce tower elements and then the optimal tower (i.e., full demonstration group), were occluded and only the end-results were observed. Duration of videos were the same for all groups, guaranteeing that all groups were exposed to tower pieces and elements for approximately the same amount of time. Children were not tested on the emulation conditions due to time and budgetary constraints. See Supplementary Material 1 for scripts and some of movies used in Experiment 2 (Movie 3.1 and 4.1).

As in Experiment 1, children saw live demonstrations; adults saw video demonstrations.

3.2.2. Open practices & data accessibility

Experiment 2 was pre-registered. The preregistration for Experiment 2 (children) can be accessed here: [https://aspredicted.org/blind.php?x=uh7p63]. Pre-registration for Experiment 2 (adults) can be found here: [https://aspredicted.org/blind.php?x=wt7fb8]. De-identified data for experiments along with the coding templates and data analysis scripts are posted at [https://osf.io/casw8/]. The materials used in these studies are widely available.

3.3. Results

3.3.1. Did summative and full imitation groups differ?

Adults' performance in Experiment 2 differed by group for the optimal base (χ^2 (2) = 18.13, p < .01, φ = 0.5) with fewer Baseline participants generating the optimal base than either imitation group (all χ^2 (2) > 8.5, all ps < 0.01, all φs > 0.34). However, in contrast to Experiment 1, there were significant group differences for the generation of the apex of the tower (χ^2 (2) = 12.078, p < .01, φ = 0.41). Specifically, only the full imitation group generated the apex more often than participants in Baseline (χ^2 (1) = 9.6, p < . 01, all φs > 0.37). No other contrast was significant. Results are summarized in Fig. 5-B.

The number of adults generating the optimal tower also differed across groups (all χ^2 (4) = 20.48, p < .01, φ = 0.53). Specifically, more participants in the social groups generated the optimal tower relative to Baseline (all χ^2 (2) > 15, all ps < 0.01, all φs > 0.46). Social groups did not differ from each other on any other measure. Results are summarized in Fig. 5-C.

As in Experiment 1, and like adults, the number of children in Experiment 2 who built the optimal base differed by group (χ^2 (2) = 26.52, p < .01, φ = 0.60): More children in the social groups produced the optimal tower base when compared to Baseline (all χ^2 (1) > 8.05, all ps < 0.01, all φs > 0.28). Social groups did not differ from each other. Results are summarized in Fig. 5-B.

The number of children who generated the optimal tower similarly differed by group (all χ^2 (2) > 23, all ps < 0.01, all φs > 0.23) with more children in the social groups producing the optimal tower than children in Baseline (all χ^2 (1) > 11, all ps < 0.01, all φs > 0.33). The differences between social groups were not significant. Results are summarized in Fig. 5-C.

As in Experiment 1, more children and adults in Baseline generated the suboptimal tower base than participants in either the summative and full imitation groups, however, these differences did not reach statistical significance after corrections (c.f., Supplementary Material 1:

Fig. 6. Experiment 2 Groups and Results for Adults. Rows correspond to (A) experimental groups, (B) proportion of subjects generating tower elements and (C) proportion of subjects generating optimal and suboptimal towers. Columns correspond to each experimental group: (1) one model summative emulation, (2) one model full emulation and (3) baseline. Note: After correction none of the measures significantly differed from Baseline.

Table A1). The fact that this trend appears in both Experiment 1 and 2 and in children as well as adults, points to a small, but nonetheless reliable, individual learning signal for this task.

3.3.2. Did summative and full emulation groups differ?

Analyses replicated those described above, but included the adult summative and full emulation groups. When compared to Baseline, there were differences in the number of individuals in each group generating the optimal base and apex (all χ^2 (4) = 24.27, p < .01, φ = 0.45) but not the suboptimal base (χ^2 (4) = 8.19, p = .09, φ = 0.26). Post-hoc tests, using the Bonferroni correction, showed that the Summative Emulation group did not significantly differ from Summative Imitation or Baseline groups for any measure (all χ^2 (1) < 5, all ps > 0.05, φ = 0.20). However, when compared to the Full Emulation group, more participants in the Summative Imitation and Full Imitation group produced the optimal base, apex and target tower (all χ^2 (1) > 5, all ps < 0.05, φ = 0.20). Emulation groups only differed when building the apex (χ^2 (1) = 5.79, p < .05, φ = 0.22). Specifically, more participants in the summative emulation groups produced the apex than those in the Full emulation group. These results suggest that, in contrast to summative imitation, there's a facilitative effect associated with combining different results from different models (i.e., without observing the corresponding actions), but not when combining different results from the *same* model. Results are summarized in Fig. 6-B, 6-C.

3.3.3. Does summative imitation produce cumulative learning?

We replicated the models used in Experiment 1 when analyzing the data of Experiment 2, except that for children this new analysis included full imitation (and excluded summative emulation); Whereas, for adults, the analysis included the same summative learning groups used in Experiment 1, in addition to full imitation and full emulation. For adults, there was a significant interaction between trial number and group (LRT: χ^2 (2) = 16.64, p < .01). Based on pairwise comparisons, this interaction was driven by the fact that whereas there were significant increases in tower height from Trial 1 to Trial 3 in the Summative Imitation (p < .01) and Baseline (p < .01) groups, the performance of the Full Imitation group was near ceiling at Trial 1, so did not change from Trial 1 to Trial 3 (p = .99).

Among children, there were significant main effects of age (LRT: χ^2 (2) = 13.87, p < .01; Supplementary Material 1: Table A3), trial (LRT: χ^2 (2) = 10.71, p < .01), and group (LRT: χ^2 (2) = 21.54, p < .01). However, in contrast to adults, the group by trial interaction was not significant (LRT: χ^2 (4) = 6.51, p = .17). Pairwise comparisons showed that 4-year old's' towers were significantly shorter than those of 5- (p < .01) and 6-year old's (p < .01). Towers in Trial 3 were taller than those in Trial 1 (p < .01) and marginally taller than those in trial 2 (p = .05). Towers in the Full Imitation group were also taller, overall, than those in Baseline (p < .01) and Summative Imitation (p < .01). No other comparison was significant. Results are summarized in Fig. 4B, Supplementary Material 1: Tables A6-A7.

3.3.4. Does summative emulation produce cumulative learning in adults?

We replicated the models used in Experiment 2 and added the summative emulation and full emulation groups. The group by trial interaction was significant (χ^2 (8) = 32.18, p < .01). This interaction was driven by the fact that whereas there were significant increases in tower height from Trial 1 to Trial 3 across most groups, including Baseline (all p < .05), the performance of the Full Imitation group was near ceiling at Trial 1, so did not change from Trial 1 to Trial 3

($p = .99$), and that of the Full emulation group was at the floor in comparisons to the other groups and remained flat ($p = .95$). Results are summarized in Fig. 4B.

3.3.5. Did adults overimitate?

The number of adults who overimitated one or more arbitrary base (atypical stacking) or apex (patterned linking) elements differed by group in Experiment 1 (Chi-square goodness of fit, Base: χ^2 (4) = 6.911, $p = .032$, $\varphi = 0.31$; Apex: χ^2 (4) = 8.117, $p = .017$, $\varphi = 0.34$) and Experiment 2 (Base: χ^2 (4) = 14.653, $p = .005$, $\varphi = 0.349$; Apex: χ^2 (4) = 16.885, $p = .002$, $\varphi = 0.375$). In Experiment 1, the number of individuals overimitating base and apex elements significantly differed from baseline (all χ^2 (3) > 5, $p < .05$, $\varphi = 0.30$). However, in Experiment 2, only participants in the full imitation group significantly overimitated both tower elements relative to those in baseline (all χ^2 (1) > 6, $p < .05$, $\varphi = 0.35$). Summative imitation groups differed from baseline in number of individuals overimitating the apex (χ^2 (1) = 8.33, $p = .016$, $\varphi = 0.42$) but not the base element (χ^2 (1) = 4.18, $p = .16$, $\varphi = 0.42$). However, summative and full imitation groups did not differ (all χ^2 (1) < 3, $p > .65$, $\varphi = 0.22$). These results show that while participants overimitated the arbitrary combination of squares to form the tower's apex above Baseline levels, they appear to have inhibited copying the atypical–and potentially suboptimal–stacking of cubes to make the tower's base. This pattern of performance is consistent with goal emulation in favor of a more stable tower base. Results are summarized in Fig. 7.

As in prior work (Berl & Hewlett, 2015; Flynn & Smith, 2012; McGuigan, 2012; McGuigan, Gladstone, & Cook, 2012; McGuigan, Makinson, & Whiten, 2011; Whiten et al., 2016), our finding of adults copying arbitrary responses when building both the base and the apex of the tower provides robust evidence of overimitation. However, participants' overimitation was not indiscriminate, which suggests that summative imitation, specifically, may act as an adaptive filter (Enquist & Ghirlanda, 2007). In particular, adults consistently copied the arbitrary combination of squares to make the tower's apex (no physical consequence) in both summative and full imitation groups. Yet, in Experiment 2, only in the full imitation group did more adults copy the atypical stacking of cubes relative to those in Baseline. Note that this response, while potentially diminishing the stability of the tower, had no impact on tower height. As such, adults' overimitation in the present task did not involve 'blanket copying' (Whiten et al., 2016), but instead, appears to have been constrained by folk physical knowledge (Lyons, Damrosch, Lin, Macris, & Keil, 2011).

3.3.6. Is summative imitation an intuitive mechanism for invention?

To address this question, data from the baseline and summative imitation groups were collapsed across Experiment 1 and 2. We first examined potential differences between adults and children in the frequency of independent invention (c.f., Table 1) for tower elements (optimal base, apex) and spontaneous combination without any demonstration. Results showed no significant difference in the rate of independent invention of the Optimal Tower between adults and children after correcting for multiple comparisons (all χ^2 (1) < 4, all $ps > 0.10$). All chi-square results are summarized in Supplementary Material 1: Table A8.

Second, we examined error responses using separate generalized linear models (GLMs) for each error type with a binomial error distribution and a logit link function to determine what predicts the probability of making (1) an error in general or (2) any of the 3 types of errors: balance, nesting, stacking (Fig. 3B-D). For each participant we coded a binary outcome variable representing the presence or absence of a given error type at any point during the experiment (i.e. the error could be made in any trial). All models included Population (children, adults), Groups (baseline, summative imitation), and the interaction of Population and Group as predictors. Baseline and summative imitation were the only two groups that were shared by both populations across experiments. However, all results for all experimental groups that included both children and adults are summarized in Fig. 8.

In terms of the probability of making any error, there was a main effect for group (Baseline > Summative Imitation, $Z = -2.40$, $p < .02$). The population by group interaction did not reach statistical significance ($Z = -1.75$, $p = .08$) but showed a tendency for children to be more likely to make an error than adults, specifically in the summative imitation group. The analysis of Stacking Errors showed a main effect for population (Children > Adults, $Z = -2.11$, $p = .04$). But neither experimental group ($Z = 0.43$, $p = .66$) nor the population by group interaction ($Z = -1.82$, $p = .07$) reached statistical significance. However, there was a non-significant tendency for children to be more likely than adults to make a stacking error in the summative imitation group in particular. The GLM for Nesting Errors produced only a significant main effect of group (Baseline > Summative Imitation, $Z = -2.12$, $p = .03$). Finally, an analysis of Balance Errors showed a main effect for population (Adults > Children, $Z = 2.14$, $p = .03$) and group (Baseline > Summative Imitation, $Z = -2.54$, $p = .01$) but no significant population by group interaction ($Z = -1.648$, $p = .10$). Full GLM results including predicted probabilities are summarized in Supplementary Material 1: Tables A9–A12.

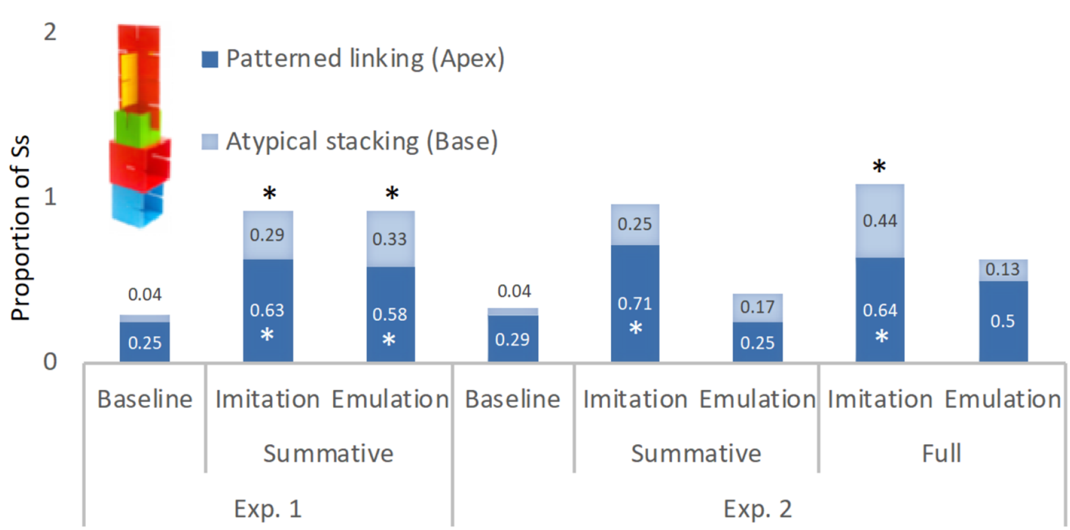

Fig. 7. Overimitation Performance. Proportion of adult participants replicating responses that were either irrelevant (patterned linking of squares) or maladaptive (atypical stacking of cubes). Note: * significantly greater than baseline.

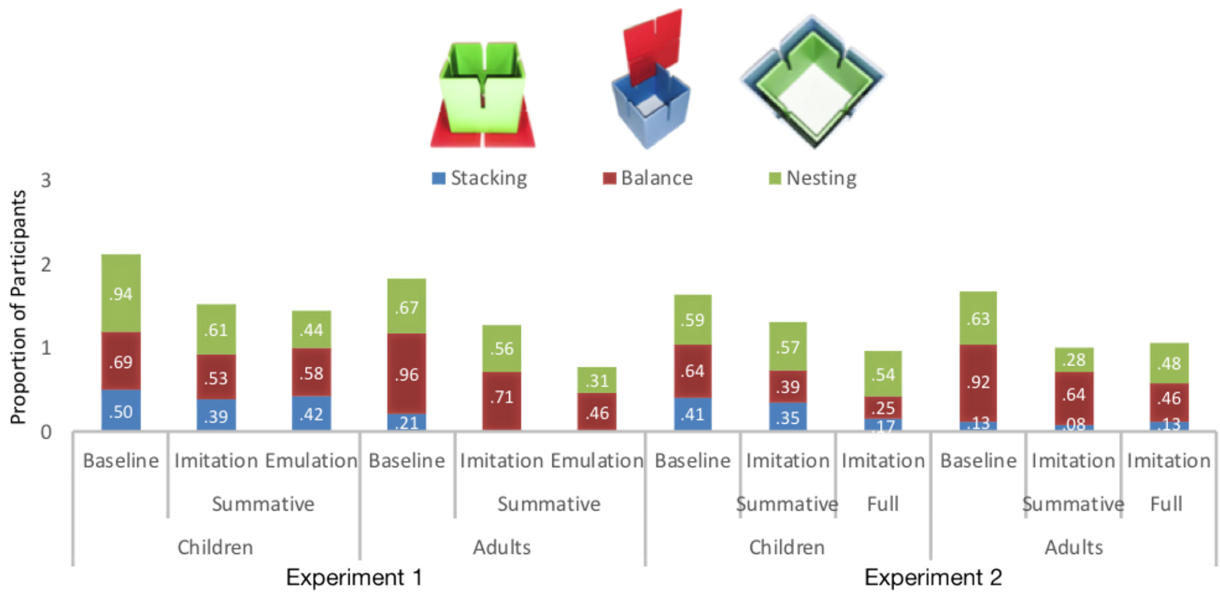

Fig. 8. Proportion of children and adults making specific types of errors in Experiments 1 and 2. Columns capture the proportion in each learning group making one of three types of errors: stacking, balance and nesting (examples of each are pictured above). Note: The maximum possible proportion of participants making a given error type = 1 (i.e., a participant that makes all three errors would have the maximum score of 3).

3.4. Discussion

When compared to independent invention (Baseline), the results of Experiment 2, involving a single model, did not differ from those of Experiment 1 involving 2 models. This similarity suggests that learning from one model involves the same skills as learning from 2 or more models. There were more potentially meaningful differences between the combination of different imitated responses (summative imitation) and the imitation of a single response (full imitation). This difference is important because summative learning whether from 1 or more models may require an asocial–undemonstrated–insight about how to optimally join base and apex. In contrast, (full) imitation is purely social and likely primes overimitation, which would inhibit such asocial insights.

In contrast to summative imitation, we found relatively poor performance in the emulation groups, particularly full emulation (c.f., Fig. 4). In fact, summative and full emulation groups did not differ from the independent invention (Baseline) group. That is, being provided with only results by a single model, did not promote invention or cumulative learning in adult participants (c.f., Lewis & Laland, 2012; Muthukrishna & Henrich, 2016 for similar results from a computational model). Together, these results provide empirical evidence to the hypothesis that under most conditions, imitation is likely to be the primary mechanism for cumulative cultural evolution.

It's unclear why summative emulation groups in Experiment 1 (2 models) differed from Baseline whereas those of Experiment 2 (1 model) did not. We had reported a similar result in an earlier study (Subiaul et al., 2015). Initially, we hypothesized that chunking (e.g., events) could explain any facilitative effect between learning different responses from one versus different models. But, if that were the case, there should have been a significant difference between all summative and full demonstration groups (whether imitation or emulation) because in summative learning groups, responses were always chunked and in the full demonstration groups they never were. While we found no difference between summative and full imitation demonstration groups, there were more consistent differences between the summative and full emulation groups, where learning was more robust in the former than the latter (c.f., Supplementary Material 1: Table A1). This pattern of performance points to the complexity of a task that superficially appears to be both familiar and simple.

Finally, we found no evidence to support the hypothesis that invention by summative imitation is itself a culturally learned skill. Although cumulative learning was more robust in adults (c.f., Fig. 3E & F), results in 10 of 11 chi-square contrasts were the same for children and adults (c.f., Supplementary Material 1: Table 1). Additionally, adults made many of the same errors as children who were more than a decade younger (c.f., Fig. 8). This was particularly true in the independent invention (Baseline) group and broadly true in the summative imitation condition where the population by group interaction failed to reach statistical significance for any of the error types. In contrast to prior studies on tool innovation (e.g., Beck et al., 2011), the present study shows that when both task difficulty and cultural knowledge are equated, children and adults are equally poor innovators.

4. General discussion & conclusions

Although 'traditions' exist in nature (Fragaszy & Perry, 2003) and some of these traditions have been described as 'cultural' (Mann, Stanton, Patterson, Bienenstock, & Singh, 2012; van Schaik et al., 2003; Whiten et al., 1999), human cultural traditions are cumulative, becoming more complex over time (Henrich, 2016; Laland, 2017; Mesoudi, 2011). Why? Here, we provide evidence that summative imitation and, to a lesser extent, emulation represent (i) spontaneous and intuitive mechanisms that children and adults alike use to innovate, as well as (ii) a means by which to aggregate adaptive behavioral responses over time. Specifically, our results show that participants who observed distinct tower elements demonstrated by one or more models, spontaneously combined these elements to produce an unobserved optimal tower. There were remarkably few significant differences between summative imitation and full imitation groups, despite the fact that the former required combining distinct behaviors in novel ways and the latter did not (c.f., Fig. 3D). However, when compared to Baseline, the differences in performance between imitation and emulation groups across experiments was notable (c.f., Figs. 3, 5-6). These results suggest that the effect of summative learning on cumulative culture is sensitive to the type or quality of information as well as the number of models available (Beck et al., 2011; Muthukrishna & Henrich, 2016).

The fact that children faithfully replicated different tower elements and adults evidenced overimitation may not be surprising to many. What should be surprising to all is that both adults and young children used their imitation learning skills to innovate in nearly identical ways. These results demonstrate that nearly two decades of higher education and direct experience balancing and stacking objects does not significantly affect invention by summative imitation in our Tower Task (Fig. 1). This outcome was neither inevitable nor artificial. First, despite our efforts to equate task difficulty and familiarity, recall that participants could make anywhere from 31 (children) to 47 (adults) different responses and over a dozen different structures. Second, given these degrees of freedom, children could have failed to evidence invention by summative learning or demonstrated adult-like performance much later in development. Third, although there was only one optimal solution, participants could have consistently favored a variety of other sub-optimal structures, including one that, although structurally unsound, was optimally tall (e.g., Fig. 3A). Finally, children and adults could have arrived at similar outcomes while differing in the types of errors they made in the process.

The parallels between the performance of children and adults, coupled with earlier work showing that summative imitation on a familiar task appears at least by age 3 (Subiaul et al., 2015), is inconsistent with the hypothesis that cumulative learning and innovation via summative learning is a culturally learned skill or 'cognitive gadget' (Heyes, 2018). Instead, these results suggest that both summative imitation and emulation represent distinct information-processing adaptations *for* cultural learning. This conclusion is strengthened by cross-cultural research showing the both the onset, fidelity and versatility of young children's imitation (Callaghan et al., 2011) and innovation (Neldner et al., 2017; Nielsen & Tomaselli, 2010) skills are similar across western and non-western cultures.

But, admittedly, such conclusion regarding the nature of innovation in humans must be tempered by the fact that the present study only included WEIRD (Western, Educated, Industrialized, Rich, and Democratic) participants, whose performance may be non-representative (Henrich, Heine, & Norenzayan, 2010). Arguably, our tower task measures a non-functional, culture-specific skill. But that assumption overlooks the fact that the production of tall, stable, structures like a tower depends on comprehending folk physical concepts such as support and balance (Kubricht, Holyoak, & Lu, 2017; Povinelli, 2012). An understanding of these folk physical concepts is inherent in structures as different as igloos, teepees and skyscrapers, but also in tools like ladders and stools as well as activities like tree-climbing. Given this, we predict little cross-cultural variation in our tower building task. Though, extending this paradigm to non-WEIRD cultures is necessary to conclusively answer this question.

Without doubt, innovations arise from multiple sources: social, asocial or, even, by accident (Kolodny et al., 2015). The question is, what is the *dominant* source and character of human innovation? Our results provide empirical support to the claim made by Muthukrishana & Henrich (2016) as well as those of Kroeber (1917) over a century ago, that—on average—*most* human innovations are inherently social; Its content is *communis generis*, of a common or shared type, rather than *sui generis* or a thing in itself. Even, when innovations arise serendipitously, the resulting products are socially-mediated (Kolodny et al., 2015; Migliano et al., 2020). For instance, in the present study, participants' optimal tower could have resulted from social processes–imitating cube stacking and square linking–as well as individual processes–inferring how best to connect these different structures. But the connection of these structures may have also been socially learned rather than individually inferred. For example, upon witnessing the joining of the flat squares (Fig. 1B-1), participants may have learned that items with ridges could be connected with other items with ridges; a type of social (emulation) learning referred to as affordance learning (Nagell, Olguin, & Tomasello, 1993) which has been implicated in other studies on innovation (Neldner et al., 2019). Alternatively, participants may have imitated the stacking of cubes and the insertion of flat squares in the ridge of *any* item with a ridge (whether another square or a cube). One or more of these explanations may account for participants' summative learning in the present study and warrants further investigation. Regardless, summative learning, whether by imitation or emulation, may represent a unique cultural learning mechanism that generates both new solutions (innovations) and products (inventions) in response to novel problems. Its primary function being the integration of outputs from one or more social (imitation or emulation) and/or asocial (inferential) learning processes. Building on the dual engine model (Legare & Nielsen, 2015), these results suggest that summative learning represents a cultural learning mechanism modulating different 'engines' (social and/or asocial) necessary for the elaboration and aggregation of knowledge.

While these results cannot entirely exclude the possibility of some asocial learning in children's and adults' intuitive inventions, various studies have shown that both children and adults turn to social, rather than individual learning, when they encounter new problems (Flynn, Turner, & Giraldeau, 2016). Research with children has also shown that neither divergent (creative) thinking nor executive functions such as inhibition or working memory significantly predict innovation. However, receptive vocabulary does (Beck et al., 2016). Beck and colleagues (Beck et al., 2016) interpreted this association between vocabulary and innovation as one between innovation and "general" intelligence. But receptive vocabularies are a measure of word learning, a cultural learning skill. If innovation depends on social learning, then language learning should predict innovativeness better than asocial cognitive processes like executive functions. This is not to say that innovation–specifically, invention–is independent of executive functions, it is not (Beaty, Silvia, Nusbaum, Jauk, & Benedek, 2014; Liu et al., 2018). Rather, executive functions (among other asocial cognitive variables) do not appear to be the principle predictor of innovation in general and invention, in particular.

While we do not expect much variation in tower-building skills cross-culturally, the effects of summative learning on cumulative learning might evidence more cross-cultural variation, at least in terms of developmental onset. Recall that young children's cumulative learning was weak relative to that of older children and adults. This result suggests that specific task-demands or experience-dependent information-processing skills are likely to constrain cumulative learning via summative learning in children.

We also do not want to overlook the role that different procedures or methodological choices might have on both children and adults' performance. In contrast to most experimental research on cumulative cultural evolution, the paradigm used here assessed cumulative learning within- not between-subjects (Caldwell & Millan, 2008). The assumption being that between-subject cumulative learning is unlikely without evidence of within-subject cumulative learning. But this assumption sidesteps an age-old problem: To what extent (or in which contexts) are macro- and micro-cumulative cultural evolutionary changes yoked (Sapir, 1917) or not (Kroeber, 1917)?

Our task also differed from those used in previous studies. While many different structures are possible with the pieces provided, optimal height and structural soundness was purposefully constrained, potentially limiting the differences between children and adults in terms of invention. In fact, there was only one possible optimal invention. Though, there were over a dozen possible structures that could be built. Nonetheless, our approach has several advantages. First, it allows us to clearly identify what was socially learned from what was individually inferred. Second, tower features are discrete and directly associated with both continuous (e.g., tower height) and categorical factors (e.g., errors) that affect structural soundness. Finally, these task features made it possible to scale the task in terms of difficulty for different populations in a predictable manner. This last point is critical to understand the differences between children and adults' potential for innovation, a major concern in prior developmental research (Beck et al., 2011; Carr et al., 2016; Neldner et al., 2017).

Although this study focuses on the artefact domain (Henrich, 2016), cumulative cultural evolution is a feature of various domains from language (Kirby, 2017) to medicinal knowledge (Migliano et al., 2020). While summative learning is unlikely to work the same across all tasks, we nonetheless expect summative learning to be a powerful source of novel knowledge (Bauer & Larkina, 2017) and cultural products (Muthukrishna, Doebeli, Chudek, & Henrich, 2018) in many different domains.[2] Consider word learning. According to the Global Language Monitor, over 5000 new words are created each year. Only about 1000 of these become widespread. Among these are blends like Brexit (i.e., British Exit [of European Union]) and compounds such as crowdfunding (i.e., raising funds from large groups). Brexit represents an example of summative emulation as both phonological and semantic features of different words are, literally, altered when combined. Crowdfunding, however, would be an example of summative imitation, as the phonological and semantic features of the two words remain, generally, unchanged when combined. Future research should explore whether summative learning similarly extends beyond the artefact domain.

In sum, the present study highlights the social nature of human creativity and innovation. The results reported here demonstrate that preschoolers and adults alike spontaneously and intuitively invent by combining the knowledge and skills of others. We refer to this way of innovating as summative learning, with summative imitation, in particular, playing a critical role. If, as we suspect, summative learning is the primary way humans innovate, it might be impossible to completely segregate cultural learning from most innovations, including transformative ones. So, rather than being *sui generis*, human inventions are, broadly, *communis generis*.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2020.104320.

Declaration of competing interest

We have no competing interests.

Acknowledgments

This study could not have been completed with the incalculable help of the following students: Leah Gillon, Maria E. Diaz, Jenna Cafora, Ilaina Feldman, Stephanie Altmann, Whitney Wright, Jasmin Uddin, Sarah Drukman, Chioma Aneke, Yuval Levin, Michael Rodriguez, Lydia Sonenklar and Courtney Sexton. We cannot thank enough the National Building Museum and the generous participation of college students, children and their parents. Finally, special thanks to Mark Nielsen, Cameron Turner, Arber Tasimi, Rachel Barr, Elizabeth Renner and three anonymous reviewers for reading earlier drafts of this manuscript. This research was funded by the George Washington University Facilitating Fund (Washington, D.C., USA) to FS.

References

Barr, R., Muentener, P., & Garcia, A. (2007). Age-related changes in deferred imitation from television by 6- to 18-month-olds. *Developmental Science, 10*(6), 910–921. https://doi.org/10.1111/j.1467-7687.2007.00641.x.

Barr, R., Muentener, P., Garcia, A., Fujimoto, M., & Chavez, V. (2007). The effect of repetition on imitation from television during infancy. *Developmental Psychobiology, 49*(2), 196–207.

Basalla, G. (1988). *The evolution of technology.* Cambridge UK: Cambridge University Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Bauer, P. J., & Larkina, M. (2017). Realizing relevance: The influence of domain-specific information on generation of new knowledge through integration in 4- to 8-year-old children. *Child Development, 88*(1), 247–262. https://doi.org/10.1111/cdev.12584.

Bauer, P. J., & San Souci, P. (2010). Going beyond the facts: Young children extend knowledge by integrating episodes. *Journal of Experimental Child Psychology, 107*(4), 452–465. https://doi.org/10.1016/j.jecp.2010.05.012.

Beaty, R. E., Silvia, P. J., Nusbaum, E. C., Jauk, E., & Benedek, M. (2014). The roles of associative and executive processes in creative cognition. *Memory & Cognition, 42*(7), 1186–1197. https://doi.org/10.3758/s13421-014-0428-8.

Beck, S. R., Apperly, I. A., Chappell, J., Guthrie, C., & Cutting, N. (2011). Making tools isn't child's play. *Cognition, 119*(2), 301–306. https://doi.org/10.1016/j.cognition.2011.01.003.

Beck, S. R., Williams, C., Cutting, N., Apperly, I. A., & Chappell, J. (2016). Individual differences in children's innovative problem-solving are not predicted by divergent thinking or executive functions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 371*(1690), https://doi.org/10.1098/rstb.2015.0190.

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), e0120180. doi:https://doi.org/10.1371/journal.pone.0120180.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences of the United States of America, 108*(Suppl. 2), 10918–10925. https://doi.org/10.1073/pnas.1100290108.

Caldwell, C. A., & Millan, A. (2008). Experimental models for testing hypotheses about cumulative cultural evolution. *Evolution and Human Behavior, 29*, 165–171. https://doi.org/10.1016/j.evolhumbehav.2007.12.001.

Caldwell, C. A., & Millen, A. E. (2008). Experimental models for testing hypotheses about cumulative cultural evolution. *Evolution and Human Behavior, 29*, 165–171. https://doi.org/10.1016/j.evolhumbehav.2007.12.001.

Caldwell, C. A., & Millen, A. E. (2009). Social learning mechanisms and cumulative cultural evolution. Is imitation necessary? *Psychological Science, 20*(12), 1478–1483. https://doi.org/10.1111/j.1467-9280.2009.02469.x.

Callaghan, T., Moll, H., Rakoczy, H., Warneken, F., Liszkowski, U., Behne, T., & Tomasello, M. (2011). Early social cognition in three cultural contexts. Monographs of the Society for Research in Child Development, 76(2), vii-viii, 1-142. doi:https://doi.org/10.1111/j.1540-5834.2011.00603.x.

Carr, K., Kendal, R. L., & Flynn, E. G. (2016). Eureka!: What is innovation, how does it develop, and who does it? *Child Development, 87*(5), 1505–1519. https://doi.org/10.1111/cdev.12549.

Choi, K., Kirkorian, H. L., & Pempek, T. A. (2018). Understanding the transfer deficit: Contextual mismatch, proactive interference, and working memory affect toddlers' video-based transfer. *Child Development, 89*(4), 1378–1393. https://doi.org/10.1111/cdev.12810.

Chomsky, N. (2015). *The minimalist program* (20th anniversary edition. ed.)Cambridge, Massachusetts; London, England: The MIT Press.

Chow, S.-C., Shao, J., Wang, H., & Lokhnygina, Y. (2017). *Sample size calculations in clinical research* (3 ed.). New York: Taylor & Francis Group.

Cosmides, L., & Tooby, J. (2002). Unraveling the enigma of human intelligence: Evolutionary psychology and the multimodular mind. In R. J. Sternberg, & J. C. Kaufman (Eds.). *The evolution of intelligence* (pp. 145–198). Mahwah, NJ: Lawrence Erlbaum Associates.

Dickerson, K., Gerhardstein, P., Zack, E., & Barr, R. (2013). Age-related changes in learning across early childhood: A new imitation task. *Developmental Psychobiology, 55*(7), 719–732. https://doi.org/10.1002/dev.21068.

Enquist, M., & Ghirlanda, S. (2007). Evolution of social learning does not explain the origin of human cumulative culture. *Journal of Theoretical Biology, 246*(1), 129–135. https://doi.org/10.1016/j.jtbi.2006.12.022.

Fay, N., De Kleine, N., Walker, B., & Caldwell, C. A. (2019). Increasing population size can inhibit cumulative cultural evolution. *Proceedings of the National Academy of Sciences of the United States of America, 116*(14), 6726–6731. https://doi.org/10.1073/pnas.1811413116.

Flynn, E., & Smith, K. (2012). Investigating the mechanisms of cultural acquisition. *Social Psychology, 43*(4), 185–195. https://doi.org/10.1027/1864-9335/a000119.

Flynn, E., Turner, C., & Giraldeau, L. A. (2016). Selectivity in social and asocial learning: Investigating the prevalence, effect and development of young children's learning preferences. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 371*(1690), https://doi.org/10.1098/rstb.2015.0189.

Fragaszy, D. M., & Perry, S. (2003). *The biology of traditions: Models and evidence.* Cambridge, UK, New York: Cambridge University Press.

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences, 9*(9), 416–423. https://doi.org/10.1016/j.tics.2005.07.004.

Heerwagen, J. H., & Orians, G. H. (2002). The ecological world of children. In J. P. H. Kahn, & S. R. Kellert (Eds.). *Children and nature: Psychological, sociocultural, and evolutionary investigations* (pp. 29–63). Cambridge: MIT Press.

Henrich, J. (Cartographer). (2016). The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter [xv, 445 pages].

Henrich, J., Boyd, R., & Richerson, P. J. (2008). Five misunderstandings about cultural evolution. *Human Nature, 19*(2), 119–137. https://doi.org/10.1007/s12110-008-9037-1.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences, 33*(2–3), 61–83. discussion 83-135 https://doi.org/10.1017/S0140525X0999152X.

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. https://doi.org/10.1016/j.cognition.2013.08.010.

Heyes, C. M. (2018). *Cognitive gadgets: The cultural evolution of thinking.* Cambridge,

[2] For related concepts in the linguistic domain, see Bauer and San Souci's (2010) 'self-derivation' paradigm, Chomsky's (2015) 'merge' and Hagoort's (2005) 'semantic unification'.

Massachusetts: The Belknap Press of Harvard University Press.
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.
Kempe, M., & Mesoudi, A. (2014). An experimental demonstration of the e ect of group size on cultural accumulation. *Evolution and Human Behavior, 35*(4), 285–290.
Kirby, S. (2017). Culture and biology in the origins of linguistic structure. *psychonomic. Bulleting and Review, 24*(1), 118–137.
Kolodny, O., Creanza, N., & Feldman, M. W. (2015). Evolution in leaps: The punctuated accumulation and loss of cultural innovations. *Proceedings of the National Academy of Sciences of the United States of America, 112*(49), E6762–E6769. https://doi.org/10.1073/pnas.1520492112.
Kroeber, A. L. (1917). The superorganic. *American Anthropologist, 19*(2), 163–213.
Kubricht, J. R., Holyoak, K. J., & Lu, H. (2017). Intuitive physics: Current research and controversies. *Trends in Cognitive Sciences, 21*(10), 749–759. https://doi.org/10.1016/j.tics.2017.06.002.
Laland, K. N. (2017). *Darwin's unfinished symphony: How culture explains the evolution of the human mind.* Princeton, NJ: Princeton University Press.
Legare, C. H., & Nielsen, M. (2015). Imitation and innovation: The dual engines of cultural learning. *Trends in Cognitive Sciences, 19*(11), 688–699. https://doi.org/10.1016/j.tics.2015.08.005.
Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). Emmeans: Estimated marginal means (aka least-squares means) (version 1.3.5.1). Retrieved from https://github.com/rvlenth/emmeans.
Lewis, H. M., & Laland, K. N. (2012). Transmission fidelity is the key to the build-up of cumulative culture. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 367*(1599), 2171–2180. https://doi.org/10.1098/rstb.2012.0119.
Liu, Z., Zhang, J., Xie, X., Rolls, E. T., Sun, J., Zhang, K., ... Feng, J. (2018). Neural and genetic determinants of creativity. *Neuroimage, 174*, 164–176. https://doi.org/10.1016/j.neuroimage.2018.02.067.
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 of London. Series 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.
Mann, J., Stanton, M. A., Patterson, E. M., Bienenstock, E. J., & Singh, L. O. (2012). Social networks reveal cultural behaviour in tool-using [corrected] dolphins. *Nature Communication, 3*(980), https://doi.org/10.1038/ncomms1983.
McGuigan, N. (2012). The role of transmission biases in the cultural diffusion of irrelevant actions. *Journal of Comparative Psychology, 126*(2), 150–160. https://doi.org/10.1037/a0025525.
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. https://doi.org/10.1371/journal.pone.0050863.
McGuigan, N., Makinson, J., & Whiten, A. (2011). From over-imitation to super-copying: Adults imitate causally irrelevant aspects of tool use with higher fidelity than young children. *British Journal of Psychology, 102*(1), 1–18. https://doi.org/10.1348/000712610X493115.
Mesoudi, A. (2011). *Cultural evolution: How Darwinian theory can explain human culture and synthesize the social sciences.* Chicago, USA: University of Chicago Press.
Mesoudi, A. (2015). Cultural evolution: A review of theory, findings and controversies. *Evolutionary Biology, 43*(4), 481–497. https://doi.org/10.1007/s11692-015-9320-0.
Migliano, A. B., Battiston, F., Viguier, S., Page, A. E., Dyble, M., Schlaepfer, R., . . . Vinicius, L. (2020). Hunter-gatherer multilevel sociality accelerates cumulative cultural evolution. Science Advances*,* 6(9), eaax5913. doi:https://doi.org/10.1126/sciadv.aax5913.
Moser, A., Zimmermann, L., Dickerson, K., Grenell, A., Barr, R., & Gerhardstein, P. (2015). They can interact, but can they learn? Toddlers' transfer learning from touchscreens and television. *Journal of Experimental Child Psychology, 137*, 137–155. https://doi.org/10.1016/j.jecp.2015.04.002.
Muthukrishna, M., Doebeli, M., Chudek, M., & Henrich, J. (2018). The cultural brain hypothesis: How culture drives brain expansion, sociality, and life history. *PLoS Computational Biology, 14*(11), e1006504. https://doi.org/10.1371/journal.pcbi.1006504.
Muthukrishna, M., & Henrich, J. (2016). Innovation in the collective brain. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 371*(1690), https://doi.org/10.1098/rstb.2015.0192.
Muthukrishna, M., Shulman, B. W., Vasilescu, V., & Henrich, J. (2014). Sociality influences cultural complexity. *Proceedings of the Biological Sciences, 281*(1774), 20132511. https://doi.org/10.1098/rspb.2013.2511.
Nagell, K., Olguin, R., & Tomasello, M. (1993). Processes of social learning in the tool use of chimpanzees (pan troglodytes) and human children (Homo sapiens). *Journal of Comparative Psychology, 107*, 174–186.
Neldner, K., Mushin, I., & Nielsen, M. (2017). Young children's tool innovation across culture: Affordance visibility matters. *Cognition, 168*, 335–343. https://doi.org/10.1016/j.cognition.2017.07.015.
Neldner, K., Redshaw, J., Murphy, S., Tomaselli, K., Davis, J., Dixson, B., & Nielsen, M. (2019). Creation across culture: Children's tool innovation is influenced by cultural and developmental factors. *Developmental Psychology, 55*(4), 877–889. https://doi.org/10.1037/dev0000672.
Nielsen, M., & Tomaselli, K. (2010). Overimitation in Kalahari Bushman children and the origins of human cultural cognition. *Psychological Science, 21*(5), 729–736. https://doi.org/10.1177/0956797610368808.
Pinker, S. (2010). Colloquium paper: The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences of the United States of America, 107*(Suppl. 2), 8993–8999. https://doi.org/10.1073/pnas.0914630107.
Povinelli, D. J. (2012). World without weight perspectives on an alien mind (pp. 1 online resource).
Price, E. E., Lambeth, S. P., Schapiro, S. J., & Whiten, A. (2009). A potent effect of observational learning on chimpanzee tool construction. *Proceedings of the Biological Sciences, 276*(1671), 3377–3383. https://doi.org/10.1098/rspb.2009.0640.
Ramsey, G., Bastian, M. L., & van Schaik, C. (2007). Animal innovation defined and operationalized. *The Behavioral and Brain Sciences, 30*(4), 393–407. discussion 407-332 https://doi.org/10.1017/S0140525X07002373.
Sapir, E. (1917). Do we need a "superorganic". *American Anthropologist, 19*(2), 441–447.
Signorella, M. L., Bigler, R. S., & Liben, L. S. (1997). A meta-analysis of children's memories for own-sex and other-sex information. *Journal of Applied Developmental Psychology, 18*(3), 429–445. https://doi.org/10.1016/S0193-3973(97)80009-3.
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*, 1410. https://doi.org/10.3389/fpsyg.2015.01410.
Team (2018). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.
Tomasello, M. (2016). Cultural learning redux. *Child Development, 87*(3), 643–653. https://doi.org/10.1111/cdev.12499.
van Schaik, C. P., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C. D., Singleton, I., et al. (2003). Orangutan cultures and the evolution of material culture. *Science, 299*(5603), 102–105. https://doi.org/10.1126/science.1078004.
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*(7), e0159920. https://doi.org/10.1371/journal.pone.0159920.
Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., & Sugiyama, Y. (1999). Cultures in chimpanzees. *Nature, 399*(6737), 682–685. https://doi.org/10.1038/21415.
Williams, H. L. (2010). Intellectual property rights and innovation: Evidence from the human genome. *Journal of Political Economy, 121*(1), 1–27.