Specialization in the **vicarious** learning of novel arbitrary sequences in humans but not orangutans

Elizabeth Renner[1,2,3*], Eric **M.** Patterson[4†], and Francys Subiaul[2,3,5]

[1]Psychology, University of Stirling, Stirling, United Kingdom, [2]Center for the Advanced Study of Human Paleobiology, The George Washington University, Washington, DC, USA, [3]Smithsonian Institution, Washington, DC, USA, [4]Biology, Georgetown University, Washington, DC, USA, [5]Department of Speech-Language-Hearing Sciences, The George Washington University, Washington, DC, USA

*Author for correspondence (lrenner@gwmail.gwu.edu).
†Present address: Office of Protected Resources, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 1315 East-West Highway, Silver Spring, MD 20910, USA

Main Text

Summary

Sequence learning underlies many uniquely human behaviors, from complex tool use to language and ritual. To understand whether this fundamental cognitive feature is uniquely derived in humans requires a comparative approach. We propose that the vicarious (but not individual) learning of novel arbitrary sequences represents a human cognitive specialization. To test this hypothesis, we compared the abilities of human **children aged 3 to 5 years** and orangutans to learn different types of arbitrary sequences (item-based and spatial-based). Sequences could be learned individually (by trial and error) or vicariously from a human (social) demonstrator or a computer (ghost control). We found that both children and orangutans recalled both types of sequence following trial-and-error learning; older children also learned both types of sequence following social and ghost demonstrations. Orangutans' success individually learning **arbitrary** sequences **shows** that their failure to do so in some vicarious learning conditions is not due to general representational problems. **These results provide** new insights into some of the most persistent discontinuities observed between humans and other great apes in terms of complex tool use, language, and ritual, all of which involve the cultural learning of novel arbitrary sequences.

1. Introduction

Sequences are pervasive features of human thoughts and actions. **As such, they** underlie many uniquely human traits, including complex tool use, language, and ritual. **Consider the sequences involved in e**veryday actions like making tea in the morning, sending a text message, and celebrating a friend's birthday. In some cases, the thoughts and corresponding actions are causally yoked (e.g., making tea). In others, they are constrained by linguistic rules and communicative norms (e.g., text versus spoken messages). Still others are governed by cultural conventions as well as idiosyncratic considerations (e.g., birthday celebrations) that are causally opaque and **"goal demoted"** (i.e., **it is difficult for an observer to discern the objective** (1–3)). Ritual sequences **include** these **last** two **features**, which **distinguish** them from instrumental sequences**, such as** tool **use,** that are causal**ly meaningful** and exhibit clear goals (1,2).

Learning novel and arbitrary sequences is not unique to humans. In fact, the learning of causally opaque, serially organized responses appears to be widely shared in the animal kingdom. Animals as different as pigeons, rats, rhesus monkeys (4), chimpanzees (5), and orangutans (6) can learn sequences of arbitrarily related items. Macaques and human adults learn sequences using the same cognitive and inferential processes (4,7). Like humans, macaques evidence increasing expertise when individually learning arbitrary sequences, demonstrating greater accuracy and more rapid acquisition with each new list as list length increases (7). They also use transitive inference (rather than associative weight) when learning the serial position of novel items in a sequence (8). However, all of these studies involved individual, direct, trial-and-error learning, not social or vicarious learning**. S**ensu **Bandura et al. (9) and** Renner et al. (10), vicarious learning happens as a result of exposure to events that are neither generated nor directly experienced by the learner, that is, usually from a conspecific or the environment.

Both monkeys and apes can **socially learn** single or familiar responses, including causally relevant sequences of familiar actions (for reviews, see (11,12)). **T**here is a growing consensus that when causally irrelevant actions are added to sequences, children faithfully copy them, while great apes omit them, copying only the causally relevant intentional actions (13–15). To date, only one study has shown **the social learning of novel sequences in primates:** rhesus monkeys, with years of expertise, successfully imitated the serial position of at least 2 (out of 4) items in a novel arbitrary sequence demonstrated by another monkey (16). The issue of expertise raises the question: is the poor performance of non-human primates in **novel social learning** tasks due to issues associated with task difficulty (e.g., encoding and recalling particular types of responses)? After all, if one cannot learn how to solve a given task via individual learning, one may also be unable to do so vicariously or socially.

This is not to say that individual and social learning are always interdependent. A growing body of research has shown that the social and individual learning of novel sequences are dissociable skills in humans (17–21). For example, various studies with preschool-aged children have now shown that the imitation of item-based sequences, involving responses to distinct items that are arbitrarily related (e.g., Ambulance → Bird → Crown; Fig. 1A), is not correlated with the imitation of similarly arbitrary spatial-based sequences (e.g., Right → Bottom → Left; Fig. 1B) within subjects (19,20). Moreover, individual differences associated with learning each sequence type by trial and error do not predict variation in learning either sequence type by imitation (20). **In other words, being a good independent learner does not necessarily make one a good imitator.** This pattern of results has led Subiaul (11,22) and colleagues (20) to hypothesize that while the individual learning of arbitrary sequences may be widely shared in the primate order, the ability to vicariously learn such sequences may be phylogenetically restricted to humans. Heyes (23) has made a similar point, arguing that "even when [animals] get the experience necessary…[they] are limited in their capacity to imitate new sequences of action" (p. 4).

What exactly is it that makes humans exceptional sequence imitators relative to non-human species? Do humans have a general facility for vicariously acquiring information from the environment, regardless of what is learned or from where? Or is this facility linked to particular content types and sources? For example, if a live model or social demonstrator were removed, but their actions and/or their effects preserved, would humans learn nonetheless? And would learning in such a condition differ from that in one involving a live agent? If humans are specialized vicarious/social learners, then learning in social conditions, as well as in a vicarious condition without a live agent (i.e., ghost control for affordance learning (24)), should be better than that in individual conditions. However, if social and asocial learning are not independent (25,26), then there may be no difference between learning in social and individual conditions. Alternatively, one might reasonably predict that performance following individual learning may be better than that following vicarious learning due to the direct experience of actions and feedback. While research has shown significant dissociations between social and individual learning in sequencing tasks in human children (19,20,27), to our knowledge there is no comparable evidence with non-human primates. Such evidence is necessary to address the question of cognitive specialization.

Here, we investigate these questions using well-established touchscreen-based sequence learning tasks. Touchscreen tasks use familiar responses in novel ways and allow for within-subject comparisons, something that is impossible to do with object-based tasks involving serial actions or events (e.g., (10,14,28)). Given that most complex responses confound seriating item-specific information (i.e., *which* objects are relevant *when* in an event) with spatial-specific information (i.e., *when* and *where* in space objects are placed), coupled with the fact that the brain independently processes *what* and *where* information (29), we employed two sequencing tasks that isolate these components. In the item-specific task (hereafter called the cognitive task), participants must select three different pictures in an item-specific order, ignoring their spatial locations. In the spatial-specific task (hereafter the spatial task),

participants must select three identical pictures in a spatial-specific order, ignoring their identity. By comparing orangutans with preschool children, our study provides unique insights into the underlying cognitive similarities and/or differences between species in sequence learning under conditions that vary in the amount and type of vicarious input.

We selected preschool-aged children (3 and 5 years old) as a comparison group for two reasons. First, we want**ed** to minimize the effect of formal education (which emphasizes complex sequence learning)**. Second,** previous studies ha**ve** shown **that** by the age of 4.5, children evidence robust imitation in both the cognitive and spatial tasks. However, younger age groups (<3.5 years) evidence a mosaic social learning pattern (e.g., copying in the cognitive but not the spatial task). Subiaul (11,22,30) has argued that the social and imitation learning skills of non-human primates may show a similar mosaic pattern.

We examine the following four hypotheses (see Table S**7** for a summary). (1**.**) Concerning how vicariously acquired information compares to individually acquired information: if humans or orangutans have a specialization for vicarious learning compared to individual learning, we would expect an advantage in performance in vicarious (Social or Ghost) over individual (Recall) learning conditions. (2**.**) Regarding the specificity of vicarious learning: if human or orangutan vicarious learning is narrowly specified for social information (obtained from agents), we would expect an advantage in performance in Social over Ghost conditions. (3**.**) Considering how task difficulty or disinterest (i.e., motivation) may confound social learning performance: if orangutans fail to perform at above-chance levels in any learning condition, we would conclude that the apes lacked the necessary motivation or the tasks were too difficult to be useful in evaluating learning competence. (4) Finally, on the relationship between social, vicarious, and individual learning in children and orangutans: if individual or vicarious learning scaffolds the development of social learning in particular tasks or conditions, **they** should predict social learning performance.

2.   Experiment 1: Children

2.1 Methods and Materials

(a) Tasks.

Two touchscreen-based tasks were used in the present study: the cognitive task and the spatial task. In the cognitive task (4,31,32), three different images appear on the screen in various locations within a 4 × 4 grid (gridlines are not visible). To solve the task, the images must be touched in a certain order governed by the contents of the pictures: for example, Ambulance → Bird → Crown (Figure 1A). After each trial, the image locations are shuffled around in the grid, but the correct order (governed by picture contents) remains the same. These rules require participants to learn the sequence based on image content rather than spatial location.

The spatial task (Figure 1B; (19,20)) is similar to the cognitive task, with the following exceptions: (i) the three picture items are identical within a trial, but change across trials; (ii) the images' locations remain fixed from trial to trial; (iii) the sequence is governed by location.

Both tasks require participants to attend to, encode, and recall different features: item identity in the cognitive task and spatial location in the spatial task. With both tasks, when there are three images, the chance of choosing the correct sequence if selecting items at random is 16.7% (that is, $1/3 \times 1/2 \times 1/1 = 0.167$).

In both tasks, the relationship between the pictures or locations themselves is arbitrary. However, when items are touched in an arbitrarily specified order, a reward (both primary and secondary reinforcers) is produced. The order of elements in a ritualized behaviour (e.g., praying to a deity) is arbitrary, in that the order itself has no clear causal connection to a (presumed) outcome. The order of elements in a linguistic utterance, by contrast, often has a clear connection to the utterance's meaning and therefore a causal connection to an outcome, even though the syntactic rules of an individual language are themselves largely arbitrary. For

example, asking someone to "**Cut the bandage**" will likely lead to a different outcome than asking them to "**Bandage the cut**". In this way, the sequences in these tasks are more like linguistic ones—with causal effects—than ritualistic ones. The causal link to a detectable outcome makes it plausible that apes would be more likely to copy sequences in these situations than in tasks examining the copying of causally irrelevant actions in sequences (13–15).

(b) Learning conditions.

Four different learning conditions were used with each task. (i) In the **Baseline** condition (individual learning), participants discovered the correct sequence independently by trial and error. (ii) The **Recall** condition (another individual learning condition) always occurred directly after the Baseline condition. Once subjects correctly entered a full three-item sequence during Baseline, there was a 30-second delay during which the computer screen was occluded. Then participants were presented with the same sequence used in the preceding Baseline condition to assess recall of the sequence. (iii) In the **Ghost** (vicarious) condition, the computer demonstrated the correct sequence three times by highlighting individual items on the screen in the target order. (iv) In the **Social** condition, a human experimenter demonstrated the correct sequence three times.

(c) Participants.

A total of 96 typically developing 3-year-old ($N = 44$; mean age in months, 41.8; SD, 5.2) and 5-year-old ($N = 52$; mean age in months, 65.3; SD, 3.8) children were recruited at a local museum or zoo following IRB-approved protocols.

(d) Apparatus.

All tasks were carried out on an iMac computer (Apple, Cupertino, CA) with a MagicTouch touchscreen panel (Keytec, Garland, TX) affixed to it. The tasks were custom-written.

(e) Procedure.

The cognitive and spatial tasks were done in blocks to avoid interference, minimize training time, and ensure understanding of the rules of the task. The order of tasks was counter-balanced; half of the children received the spatial block first and the other half received the cognitive block first. A brief training phase preceded the first experimental condition for each task. After training, the testing phase began. The conditions within each task were counterbalanced (but Baseline was always immediately followed by Recall). Four different conditions (delineated above), and a total of three different sequences, were used for each task, for a 2 (tasks) × 4 (learning conditions) design. A sequence consisted of a set of three images (for the cognitive task) or three locations (for the spatial task). A trial consisted of an attempt (either correct or incorrect) to enter a sequence. After an incorrect response on a trial, a new trial with the same sequence started; in the cognitive task, the spatial locations of the pictures were shuffled, and in the spatial task, the identity of the pictures changed.

For the Baseline and Recall conditions, performance from the first presentation of a sequence (first trial, T1), as well as subsequent trials until the correct sequence was entered, was measured. Once children selected the items in the correct order in one condition, testing was complete in that condition, and a new condition was begun. For additional details on the methods, see the Supplementary Material.

(f) Statistical analysis.

Two measures were used to quantify performance. Performance on the first trial (T1) of each condition was a binomial measure, with values of correct (1) or incorrect (0). T1 is the

strictest measure of learning. In conditions with demonstrations, T1 represents a **"pure"** measure of vicarious learning because success on any trials after the first trial could be due to vicarious learning, individual learning, or a combination of both. The only way to perform at above-chance levels on this measure is to have prior knowledge of the sequence (by individual learning or vicarious learning). However, T1 cannot give information about what happens after the first trial. An individual who first enters a correct sequence on trial 2 and one who first enters a correct sequence on trial 10 receive the same score on the T1 measure (both score a 0). As such, T1 excludes partial learning.

A second measure, the correct:incorrect (CI) presses measure, is a per-trial measurement of performance. It consists of two response variables: out of the first two presses in a trial, the number of correct selections and the number of incorrect selections. A fully correct trial (Picture 1 → Picture 2 → Picture 3) results in values of [2:0]; a partially correct trial (Picture 1 → Picture 3) results in values of [1:1]; and an incorrect trial (Picture 2 or Picture 3 selected first) results in values of [0:2]. The CI measure can show how quickly an individual finds a solution if they do not do so on the first trial, and also how performance changes within a condition. For children, CI values for each trial until the first correct trial were calculated for each condition; that is, if an individual first entered 2 incorrect trials and then 1 correct trial, a total of 3 CI values were calculated. The first 4 trials for each condition were included in the statistical models, to maintain consistency with the orangutan data structure (see Section 3.1 below); if a child entered the correct sequence before they had performed 4 trials, the number of CI values corresponded to the total number of trials performed by the child.

For the main statistical analyses, generalized linear mixed models (GLMMs) **implemented in *lme4* in R (33)** were used to examine the fixed effects of condition, age group (3- and 5-year-olds), and their interactions. **We analyzed the same data with Markov chain Monte Carlo GLMMs in a Bayesian framework using the *MCMCglmm* package (34), to determine if analysis method affected the results (see Supplementary Material Section**

1.7). It did not; therefore, we report here the results of the *lme4* analyses. Regression

analyses (general linear models or GLMs) were used to evaluate the degree to which learning in

certain conditions predicted that in others.

2.2 Results

(a) Cognitive task performance by condition and age.

We used a binomial GLMM with a response variable of CI; fixed effects of age group

(two levels: 3 years and 5 years), condition (four levels: Baseline, Ghost, Social, Recall), and

their interaction; and a random effect of participant (ID). There was no main effect of age group,

but there was a main effect of condition and an interaction between condition and age group. To

explore this interaction, we separated the data by age and re-ran the models with condition as

the only fixed effect. For 3-year-olds, pairwise contrasts between conditions (using Tukey's

correction for multiple comparisons) indicated that performance was significantly better in the

Social condition than Baseline ($b = 0.68$, SE $= 0.23$, Z $= 3.0$, $p = 0.013$); no other pairwise

contrast showed a significant difference (all ps > 0.2; Figure 2), including comparisons between

Social and Recall and between Social and Ghost. For 5-year-olds, pairwise contrasts between

conditions indicated that when compared to Baseline, performance was significantly better in

the Social ($b = 1.9$, SE $= 0.27$, Z $= 7.2$, $p < 0.001$), Ghost ($b = 0.64$, SE $= 0.20$, Z $= 3.2$, $p =$

0.0077), and Recall ($b = 1.1$, SE $= 0.21$, Z $= 5.1$, $p < 0.001$) conditions. Additionally,

performance in the Social condition was better than that in the Recall ($b = 0.81$, SE $= 0.28$, Z $=$

2.9, $p = 0.017$) and Ghost ($b = 1.3$, SE $= 0.27$, Z $= 4.6$, $p < 0.001$) conditions.

(b) Spatial task performance by condition and age.

We used an analysis analogous to that described above and found that, similarly to the

cognitive task, there was no main effect of age group, but there was a main effect of condition

and an interaction between condition and age group. We again separated the data by age to explore this interaction, and ran models with condition as the only fixed effect. For 3-year-olds, pairwise contrasts between conditions (with Tukey's correction) indicated that performance was significantly better in the Recall condition than in Baseline ($b = 0.61$, SE = 0.21, Z = 2.9, $p = 0.018$); no other pairwise contrast showed a significant difference (all ps > 0.1; Figure 3). For 5-year-olds, pairwise contrasts between conditions indicated that when compared to Baseline, performance was significantly better in the Social ($b = 0.95$, SE = 0.22, Z = 4.3, $p < 0.001$), Ghost ($b = 0.89$, SE = 0.20, Z = 4.4, $p < 0.001$), and Recall ($b = 1.2$, SE = 0.21, Z = 5.6, $p < 0.001$) conditions. In contrast to the cognitive task described above, however, no other pairwise contrast showed a significant difference (all ps > 0.6), including the comparisons between Social and Recall and between Social and Ghost.

For results of the analyses of both tasks with the T1 measure, see the Supplementary Material.

(c) Differences between tasks.

We examined whether children's overall performance was better in the spatial or the cognitive task by using a binomial GLMM with a response variable of CI, a fixed effect of task (**two** levels: cognitive and spatial), and a random effect of participant. There was no main effect of task ($b = -0.08$, SE = 0.073, Z = -1.2, $p = 0.25$), indicating that children's performance overall was not better or worse in either task.

(d) Relationships between performance in Social and other conditions.

We examined whether performance in the various conditions predicted children's social learning performance in each task using GLMs (see Supplementary Material **Section 1.6** for details). For the cognitive task, while age group (5-year-olds > 3-year-olds; $b = -0.78$, SE = 0.26, Z = -3.0, $p = 0.0025$) was a significant predictor of performance in the cognitive Social

condition, performance in the other conditions (cognitive Ghost, cognitive Recall, spatial Ghost, spatial Recall, **and** spatial Social) was not (all ps > 0.08). For the spatial task, age group (5-year-olds > 3-year-olds; $b = -0.40$, SE = 0.16, Z = -2.5, $p = 0.014$) was a significant predictor of performance in the spatial Social condition, as was spatial Ghost **performance** ($b = 0.54$, SE = 0.19, Z = 2.8, $p = 0.005$; all other ps > 0.07). Results are summarized in Supplementary Figure S3.

3.   Experiment 2: Orangutans

3.1 Methods and Materials

(a) Tasks.

The cognitive and spatial tasks described above were used to test the orangutans.

(b) Conditions.

The same four learning conditions described above for children were used.

(c) Participants.

Three adult orangutans living at Smithsonian's National Zoo in Washington, DC, participated in this study. Demographic details of the orangutans are shown in Table S**3** in the Supplementary Material. The protocols for this study were approved by the Institutional Animal Care and Use Committees of the George Washington University and the Smithsonian Institution.

(d) Apparatus.

The apparatus used for orangutans was similar to the one used for children, and was affixed to a mobile cart that allowed testing in the orangutans' living enclosures. During testing,

the screen was placed against the enclosure mesh so that an orangutan could interact with the touchscreen and access rewards (grapes) delivered through a feeding tube.

(e) Procedure.

Training and testing on the cognitive and spatial tasks were done in blocks to minimize training time and between-task interference, and to maximize understanding of the task rules. Iris received the cognitive block first, while Batang and Kyle received the spatial block first. For additional details, see the Supplementary Material.

(i) Training. In the training sessions, orangutans were given three demonstrations of a two-item sequence by a familiar zookeeper. They then had 4 consecutive opportunities (trials) per sequence to enter the correct two-item sequence, and received rewards for correct performance. They saw four different sequences per training session. Upon reaching a performance criterion, they moved to the testing phase.

(ii) Testing. Experimental trials used three-item sequences. Conditions within each experimental block were counterbalanced, so that each orangutan received the Ghost condition first in one block and the Social condition first in the other block. Each condition **block** consisted of 12 sessions; in addition, 12 sessions that comprised both Baseline and Recall conditions were interspersed throughout the Ghost and Social sessions, not performed in a block, to distribute any potential changes in expertise.

During testing, as during training, orangutans were given four trials per sequence in the Recall, Ghost, and Social conditions. The number of response trials was limited in order to ensure that session length was predictable and motivation to respond correctly was high (i.e., there were relatively few opportunities for a reward). Orangutans were given up to 35 trials in the Baseline condition to discover the correct sequence initially. If they did not do so, no Recall condition was begun; instead, the next sequence in a session was begun.

(f) Statistical analysis.

For orangutans, both T1 and CI values were used. CI values were calculated for trials 1 to 4 in the Baseline, Recall, Ghost, and Social conditions.

GLMMs **implemented in *lme4* in R** were used to examine the fixed effect of condition. **As for the children's data, we also created GLMMs for the orangutan data using the *MCMCglmm* package (34); see Supplementary Material Section 2.9. The method of analysis did not affect the results, so we report here the results of the *lme4* analyses. T**o evaluate the degree to which performance in **other** conditions predicted that in the Social conditions**, we used *MCMCglmm* because the *lme4* models did not converge.**

3.2 Results

(a) Cognitive task performance by condition.

We used a binomial GLMM, with a response variable of CI, a fixed effect of condition (four levels: Baseline, Ghost, Social, Recall), and random effects of participant and demonstrator. There was a main effect of condition; pairwise contrasts between conditions (using Tukey's correction) indicated that performance was significantly better in the Recall condition than in the Baseline ($b = 0.35$, SE $= 0.096$, Z $= 3.7$, $p = 0.0013$), Social ($b = 0.27$, SE $= 0.087$, Z $= 3.2$, $p = 0.0086$), and Ghost ($b = 0.41$, SE $= 0.090$, Z $= 4.6$, $p < 0.001$) conditions. No other contrasts were significantly different. Results are summarized in Figure 2 and Figure S4 in the Supplementary Material.

(b) Spatial task performance by condition.

We repeated the analysis described above **for the spatial task** and found a main effect of condition. Pairwise contrasts between conditions (using Tukey's correction) indicated that

when compared to Baseline, performance was significantly better in the Recall ($b = 1.0$, SE = 0.098, $Z = 10$, $p < 0.001$), Social ($b = 0.45$, SE = 0.10, $Z = 4.5$, $p < 0.001$), and Ghost ($b = 0.37$, SE = 0.10, $Z = 3.7$, $p = 0.0013$) conditions. Additionally, performance in the Recall condition was better than that in both the Social ($b = 0.55$, SE = 0.091, $Z = 6.0$, $p < 0.001$) and the Ghost ($b = 0.63$, SE = 0.089, $Z = 7.1$, $p < 0.001$) conditions. There were no other significant differences, including between the Social and Ghost conditions. Results are summarized in Figure 3 and Figure S5 in the Supplementary Material.

(c)   Differences between tasks.

We examined whether orangutans' overall performance was better in the spatial or the cognitive task by using a binomial GLMM with a response variable of CI, a fixed effect of task, and a random effect of participant. There was a main effect of task ($b = 0.28$, SE = 0.045, $Z = 6.2$, $p < 0.001$), indicating that orangutans performed better in the spatial than the cognitive task.

(d)   Relationships between performance in the Social and other conditions.

We examined whether performance in the various conditions predicted orangutans' social learning performance in each task **using *MCMCglmm* (see Supplementary Material Section 2.8)**. For the cognitive task, **the model** indicated that performance in the other conditions (cognitive Ghost, cognitive Recall, spatial Ghost, spatial Recall, and spatial Social) did not significantly predict performance in the cognitive Social condition (all ps > 0.**14**). For the spatial task, performance in **the** other conditions **(spatial Ghost, spatial Recall, cognitive Ghost, cognitive Recall, and cognitive Social) did not** significantly predict performance in the spatial Social condition (all ps > 0.06). **One condition,** cognitive Social, **had a marginal** negative relationship **with spatial Social performance (see Supplementary Table S4). S**ee Figure S3 **for a summary**.

(e) Comparing Recall performance with chance.

To compare the performance of orangutans with chance levels, we performed chi-square tests on the distribution of first-trial (T1) responses (correctly pressing all three items [ABC], correctly pressing the first but not the second item [AC], or pressing an initial incorrect item [B or C]) for each individual. The full results are reported in Table S**6** in the Supplementary Material. For the Recall conditions, Batang's responses in both the cognitive and spatial tasks included more ABC responses than expected by chance; Iris's responses in the spatial task (but not the cognitive task) included more ABC responses than expected by chance; and Kyle's responses in the cognitive task included fewer ABC responses than expected by chance. Kyle's responses in the Ghost condition of the spatial task also included fewer ABC responses than expected by chance. No other result differed significantly from chance.

4.   Discussion

Several uniquely human traits, such as language and ritual, involve arbitrary sequences that are culturally rather than causally specified, and vicariously rather than individually learned. The ubiquity of both language and ritual in human activities prompts the question: *Do humans have a specialization for vicariously learning arbitrary sequences across tasks and problem domains?* Consider that ritual, for example, includes arbitrary sequences of actions that are causally opaque and goal demoted (35). Language—syntax specifically—similarly consists of arbitrary sequences of words, but in contrast to ritual, words and phrases can be clearly linked to causal outcomes (36). In addition to their sequential features, the abstract and symbolic nature of the tokens (i.e., ritual acts and words)—representing unobservable concepts or absent entities—used in both domains links the evolution of ritual and language (37,38).

Here we present evidence that humans at an early age, **but not orangutans,** possess a **specialization** for **vicarious**ly learning some arbitrary sequences (Table S**7**). In each task, **participants** received individual (Recall) or vicarious information either from a social agent serving as the model (Social) or provided only by a computer, an artificial agent (Ghost). **The performance of orangutans and young children on the tasks diverged.** Regardless of task, orangutans learned sequences best in the Recall condition. They also evidenced **some** limited vicarious learning in the spatial task, with performance following the pattern of *Recall > Social ≈ Ghost > Baseline*. Such results show that the orangutans had both the motivation and ability to encode and recall novel arbitrary sequences in individual learning conditions, and in some vicarious conditions. In contrast, 5-year-old children showed proficient learning in all individual and vicarious conditions, regardless of task. Like orangutans, 3-year-old children's performance in the spatial task was best in the Recall condition. However, unlike orangutans, in the cognitive task their performance was best in the Social condition.

To our surprise, orangutans appeared to learn spatial-based sequences (spatial task) better than item-based sequences (cognitive task), despite the fact that they had less expertise in the former than the latter. This pattern differs from that observed in monkeys (who evidenced social learning in the cognitive task **(16)**) and the developmental pattern in children observed here and in other studies (19,20). There are several possible explanations for this result. First, there is empirical evidence that captive (39) as well as wild (40,41) orangutans have excellent spatial memory, as evidenced by the ability to form cognitive maps of complex habitats, which includes avoiding previously depleted sites in experimental tasks and revisiting preferred sites in the wild. Second, in search tasks, orangutans, like all other non-human great apes (gorillas, chimpanzees, and bonobos) and 1-year-old infants, favored the use of a spatial rather than a feature-based memory strategy, while 3-year-olds showed the reverse strategy (42). Finally, in a previous study by Swartz and colleagues (43), the orangutans (one of which was involved in this study) spontaneously used a spatial strategy (selecting items from right to left) when encoding

and recalling unordered items on a touchscreen task similar to the cognitive task used here. These factors may explain orangutans' comparatively better performance in the spatial than the cognitive task.

Children did not show any overall performance differences between the tasks, and the pattern of their performance differences by age replicates and extends previously reported results (19,20,44,45).

Species differences are further highlighted by our **predictive** analyses (Table**s S1, S2, S4, and S5** and Figure S3 in the Supplementary Material). These show that children's spatial Social performance is predicted by their spatial Ghost performance (consistent with children **being** adept vicarious learners, regardless of source). However, **neither** orangutans' spatial **Social performance nor their cognitive** Social performance is **significantly** predicted by **either** their individual learning (Recall) **or their vicarious learning (Ghost). Neither** orangutans' **nor** children's Social performance across tasks was predicted by their Recall performance. Together these results suggest that social and individual learning may be dissociable in **both** humans and **orangutans**, consistent with previous studies (19,20) and theories of a mosaic architecture of social learning (11,22,30).

Some limitations of the present study should be considered for designing future research. In addition to testing other great apes, it would be useful to expose participants to incorrect as well as correct responses. Errors, executed by conspecifics, have been associated with more robust social learning in both children (19,20,46) and monkeys (47,48). This would reveal whether the pattern of results reported here would change if apes and children were provided with models that showed both correct and incorrect responses. Additionally, the copying of non-arbitrary causal sequences by orangutans should be tested, for example in physical tasks that visibly require certain orders of actions. This would indicate whether it is the (arbitrary) relationship between elements that makes it difficult for orangutans to vicariously learn sequences in these tasks.

In sum, these results show that humans, from an early age, have a facility to learn novel arbitrary sequences from others in a way that orangutans do not. Is this due to the fact that humans are exposed to more (and perhaps more dependent on) sequences of actions than orangutans? Consider that wild orangutans perform some serial actions such as the daily construction of their nests for night-time sleep. This involves selecting a site for the nest, making a foundation of larger branches, and sometimes adding embellishments (49). While the seriation of some of these actions is instrumental (e.g., a nest could not be constructed in a different order), other behaviors appear to have less of an instrumental role (e.g., adding a "rim" around the edge of the nest, or other embellishments called "artistic" features (50,51)). **While** wild orangutans may sometimes be exposed to and use sequences in cases like this, sequences are ubiquitous in children's lives. In fact, from the moment children wake up in the morning to when their heads are placed on a pillow at night, children's days are organized into a series of elaborate routines that include hierarchically organized sequences. Might such experiences explain the species differences observed here? Or do humans rely on and use such elaborate sequences and routines because learning them comes so easily and naturally?

While most developmental research has focused on the unique pressures faced by human children to learn new instrumental skills by imitation (52–54), less attention has been paid to the challenges associated with **the vicarious and social** learning **of** arbitrary sequences, critical for both language and ritual (55,56). The evidence that does exist suggests that placing serial responses in a "ritual" context enhances imitation fidelity (56,57). For example, making sequences causally opaque and without an obvious end goal—two core features of rituals (35)—increases, rather than decreases, imitation fidelity. But the fact is that children in general excel at imitating all types of sequences (58–60). While our results confirm that humans and orangutans share various individual sequence learning skills, the faithful copying of socially demonstrated arbitrary sequences is highly developed in humans relative to orangutans. This is consistent with the hypothesis that imitation of novel arbitrary sequences is

a human cognitive specialization (22,61). From a developmental perspective, we do not have enough information to determine the extent to which this specialization is acquired via experience (26). And from an evolutionary perspective, we cannot say conclusively whether this skill precipitated complex tool use, language, and symbolic rituals; whether an increasing dependence on these skills placed unique pressures on the ability to vicariously learn novel arbitrary sequences; or even whether these suites of skills coevolved (23,37,57,62,63). Regardless, the interdependence between the ability to vicariously learn sequences and these uniquely human behaviors is unmistakable.

5. Conclusion

Do humans possess a cognitive specialization for vicariously learning novel sequences? The evidence presented here is consistent with the specialization hypothesis showing that at a young age, children are particularly adept, when compared to orangutans, at faithfully imitating arbitrary sequences. Orangutans' comparatively poor performance on the same tasks and conditions cannot be explained by some general representational deficit or a lack of interest (or motivation), as they showed significant learning in individual conditions and even some vicarious conditions. However, differences may be explained by the fact that vicarious sequence learning underlies many uniquely human behaviors that range from complex tool use to language and ritual. These results raise the question: is the relative facility by which humans vicariously learn novel sequences a cause of the emergence of ritual and language or, as Heyes (26) has suggested, is the specialization a product of these cultural activities? We may never know for sure. What we can say, however, is that the few apes that have been raised in human homes or given language training do not show the same facility for learning complex sequences as young human children, whether imitating novel actions on objects (22) or the sequencing of signs to communicate (64).

Acknowledgments

References

1. Kapitány R, Nielsen M. Adopting the ritual stance: The role of opacity and context in ritual and everyday actions. Cognition. 2015;145:13–29.
2. Kapitány R, Nielsen M. The ritual stance and the precaution system: the role of goal-demotion and opacity in ritual and everyday actions. Relig Brain Behav. 2017;7(1):27–42.
3. Kapitány R, Nielsen M. Ritualized objects: How we perceive and respond to causally opaque and goal demoted action. J Cogn Cult. 2019;19(1–2):170–94.
4. Terrace HS. The simultaneous chain: A new approach to serial learning. Trends Cogn Sci. 2005;9(4):202–10.
5. Inoue S, Matsuzawa T. Working memory of numerals in chimpanzees. Curr Biol. 2007;17(23):R1004–5.
6. Renner E, Price EE, Subiaul F. Sequential recall of meaningful and arbitrary sequences by orangutans and human children: Does content matter? Anim Cogn. 2016;19:39–52.
7. Terrace HS, Son LK, Brannon EM. Serial expertise of rhesus macaques. Psychol Sci. 2003;14(1):66–73.
8. Merritt DJ, Terrace HS. Mechanisms of inferential order judgments in humans (*Homo sapiens*) and rhesus monkeys (*Macaca mulatta*). J Comp Psychol. 2011;125(2):227–38.
9. **Bandura A, Ross D, Ross SA. Vicarious reinforcement and imitative learning. J Abnorm Soc Psychol. 1963;67(6):601–7.**
10. Renner E, Atkinson M, Caldwell CA. Squirrel monkey responses to information from social demonstration and individual exploration using touchscreen and object choice tasks. PeerJ. 2019;7:e7960.
11. Subiaul F. The imitation faculty in monkeys: evaluating its features, distribution and evolution. J Anthropol Sci. 2007;85:35–62.
12. Whiten A, McGuigan N, Marshall-Pescini S, Hopper LM. Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. Philos Trans R Soc London B Biol Sci. 2009;364(1528):2417–28. Available from: http://rstb.royalsocietypublishing.org/content/364/1528/2417.abstract
13. Clay Z, Tennie C. Is overimitation a uniquely human phenomenon? Insights from human children as compared to bonobos. Child Dev. 2018;89(5):1535–44.
14. Horner V, Whiten A. Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). Anim Cogn. 2005;8(3):164–81.
15. Nielsen M, Susianto EWE. Failure to find over-imitation in captive orangutans (*Pongo pygmaeus*): implications for our understanding of cross-generation information transfer. In: Håkansson J, editor. Developmental psychology. New York, NY: Nova Science Publishers; 2010. p. 153–67.
16. Subiaul F, Cantlon JF, Holloway RL, Terrace HS. Cognitive imitation in rhesus macaques. Science. 2004;305:407–10.
17. Boutin A, Fries U, Panzer S, Shea CH, Blandin Y. Role of action observation and action in sequence learning and coding. Acta Psychol (Amst). 2010;135(2):240–51. Available from: http://dx.doi.org/10.1016/j.actpsy.2010.07.005
18. Boutin A, Badets A, Salesse RN, Fries U, Panzer S, Blandin Y. Practice makes transfer of motor skills imperfect. Psychol Res. 2012;76(5):611–25.
19. Subiaul F, Anderson S, Brandt J, Elkins J. Multiple imitation mechanisms in children. Dev Psychol. 2012;48(4):1165–79.
20. Subiaul F, Patterson EM, Schilder B, Renner E, Barr R. Becoming a high-fidelity - *super* - imitator: what are the contributions of social and individual learning? Dev Sci. 2015;18(6):1025–35.
21. Subiaul F, Zimmermann L, Renner E, Schilder B, Barr R. Defining elemental imitation

mechanisms: a comparison of cognitive and motor-spatial imitation learning across object- and computer-based tasks. J Cogn Dev. 2016;17(2):221–43. Available from: http://dx.doi.org/10.1080/15248372.2015.1053483

22. Subiaul F. What's special about human imitation? A comparison with enculturated apes. Behav Sci (Basel). 2016;6:13. Available from: http://www.mdpi.com/2076-328X/6/3/13

23. Heyes C. Homo imitans? Seven reasons why imitation couldn't possibly be associative. Philos Trans R Soc B Biol Sci. 2016;371:20150069. Available from: http://rstb.royalsocietypublishing.org/lookup/doi/10.1098/rstb.2015.0069

24. Whiten A, McGuigan N, Marshall-Pescini S, Hopper LM. Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. Philos Trans R Soc B Biol Sci. 2009;364(1528):2417–28.

25. Heyes C. What's social about social learning? J Comp Psychol. 2012;126(2):193–202. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21895355

26. Heyes C. Cognitive gadgets: the cultural evolution of thinking. Cambridge, MA: The Belknap Press of Harvard University Press; 2018.

27. Subiaul F, Patterson EM, Zimmermann L, Barr R. Only domain-specific imitation practice makes imitation perfect. J Exp Child Psychol. 2019;177:248–64. Available from: https://doi.org/10.1016/j.jecp.2018.07.004

28. Bauer PJ, Mandler JM. Putting the horse before the cart: the use of temporal order in recall of events by one-year-old children. Dev Psychol. 1992;28(3):441–52.

29. Mishkin M, Ungerleider LG. Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. Behav Brain Res. 1982;6(1):57–77.

30. Subiaul F. Dissecting the imitation faculty: the multiple imitation mechanisms (MIM) hypothesis. Behav Processes. 2010;83:222–34.

31. Terrace HS. Chunking and serially organized behavior in pigeons, monkeys and humans. In: Cook RG, editor. Avian visual cognition [On-line]. Medford, MA: Comparative Cognition Press; 2001.

32. Terrace HS. Comparative psychology of chunking. In: Fountain S, editor. Animal cognition and sequential behavior. Dordrecht, Netherlands: Kluwer Academic Publishing; 2001. p. 23–56.

33. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017. Available from: https://www.r-project.org/

34. Hadfield JD. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. J Stat Softw. 2010;33(2):1–22.

35. Nielsen M, Tomaselli K, Kapitány R. The influence of goal demotion on children's reproduction of ritual behavior. Evol Hum Behav. 2018;39(3):343–8.

36. Pinker S. The language instinct. New York, NY: Harper Perennial Modern Classics; 2007.

37. Deacon TW. The symbolic species: the co-evolution of language and the brain. 1st ed. New York: W. W. Norton; 1997.

38. Donald M. Origins of the modern mind: three stages in the evolution of culture and cognition. Cambridge, MA: Harvard University Press; 1991.

39. MacDonald SE, Agnes MM. Orangutan (*Pongo pygmaeus abelii*) spatial memory and behavior in a foraging task. J Comp Psychol. 1999;113(2):213–7.

40. Vogel ER, Haag L, Wich SA, Bastian ML, van Schaik CP. Factors affecting foraging decisions in a wild population of sympatric orangutans (*Pongo pygmaeus wurmbii*) and white-bearded gibbons (*Hylobates albibarbis*): evidence of cognitive maps. In: American Journal of Physical Anthropology 77th Annual Meeting. 2008. p. S215.

41. Markham KE, van Noordwijk MA, Vogel ER. Factors influencing revisitation rates to feeding trees in wild Bornean orangutans (*Pongo pygmaeus wurmbii*) in Central Kalimantan, Indonesia. In: American Journal of Primatology 37th Annual Meeting. 2014. p. 75.

42. Haun DBM, Call J, Janzen G, Levinson SC. Evolutionary psychology of spatial representations in the Hominidae. Curr Biol. 2006;16(17):1736–40.

43. Swartz KB, Himmanen SA, Shumaker RW. Response strategies in list learning by orangutans (*Pongo pygmaeus* × *P. abelii*). J Comp Psychol. 2007;121(3):260–9.

44. Subiaul F, Lurie H, Romansky K, Klein T, Holmes D, Terrace H. Cognitive imitation in typically-developing 3- and 4-year olds and individuals with autism. Cogn Dev. 2007;22:230–43.

45. Subiaul F, Vonk J, Rutherford MD. The ghosts in the computer: the role of agency and animacy attributions in "ghost controls." PLoS One. 2011;6(11):e26429.

46. Want SC, Harris PL. Learning from other people's mistakes: causal understanding in learning to use a tool. Child Dev. 2001;72(2):431–43.

47. Ferrucci L, Nougaret S, Genovesio A. Macaque monkeys learn by observation in the ghost display condition in the object-in-place task with differential reward to the observer. Sci Rep. 2019;9(1):401.

48. Monfardini E, Hadj-Bouziane F, Meunier M. Model-observer similarity, error modeling and social learning in rhesus macaques. PLoS One. 2014;9(2):e89825. Available from: http://dx.plos.org/10.1371/journal.pone.0089825

49. Russon AE, Handayani DP, Kuncoro P, Ferisa A. Orangutan leaf-carrying for nest-building: toward unraveling cultural processes. Anim Cogn. 2007;10(2):189–202.

50. van Schaik CP, Ancrenaz M, Borgen G, Galdikas B, Knott CD, Singleton I, et al. Orangutan cultures and the evolution of material culture. Science. 2003;299:102–5.

51. Bastian ML, Van Noordwijk MA, Van Schaik CP. Innovative behaviors in wild Bornean orangutans revealed by targeted population comparison. Behaviour. 2012;149(3–4):275–97.

52. Csibra G, Gergely G. Natural pedagogy. Trends Cogn Sci. 2009;13(4):148–53.

53. Lyons DE, Damrosch DH, Lin JK, Macris DM, Keil FC. The scope and limits of overimitation in the transmission of artefact culture. Philos Trans R Soc London B Biol Sci. 2011;366(1567):1158–67.

54. Tomasello M. Cultural learning redux. Child Dev. 2016;87(3):643–53.

55. Herrmann PA, Legare CH, Harris PL, Whitehouse H. Stick to the script: the effect of witnessing multiple actors on children's imitation. Cognition. 2013;129(3):536–43. Available from: http://dx.doi.org/10.1016/j.cognition.2013.08.010

56. Legare CH, Wen NJ, Herrmann PA, Whitehouse H. Imitative flexibility and the development of cultural learning. Cognition. 2015;142:351–61.

57. Legare CH, Nielsen M. Imitation and innovation: the dual engines of cultural learning. Trends Cogn Sci. 2015;19(11):688–99. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1364661315001825

58. Bauer PJ. Holding it all together: how enabling relations facilitate young children's event recall. Cogn Dev. 1992;7(1):1–28.

59. Loucks J, Meltzoff AN. Goals influence memory and imitation for dynamic human action in 36-month-old children. Scand J Psychol. 2013;54(1):41–50.

60. Loucks J, Mutschler C, Meltzoff AN. Children's representation and imitation of events: how goal organization influences 3-year-old children's memory for action sequences. Cogn Sci. 2017;41(7):1904–33.

61. Sherwood CC, Subiaul F, Zawidzki TW. A natural history of the human mind: tracing evolutionary changes in brain and cognition. J Anat. 2008;212(4):426–54.

62. Henrich J. The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter. Princeton: Princeton University Press; 2016.

63. Renner E, Zawidzki T. Minimal cognitive preconditions on the ratchet. In: Di Paolo LD, Di Vicenzo F, De Petrillo F, editors. Evolution of primate cocial cognition: interdisciplinary

evolution research. Springer International Publishing AG; 2018. p. 249–65.

64. Lyn H, Greenfield PM, Savage-Rumbaugh S, Gillespie-Lynch K, Hopkins WD. Nonhuman primates do declare! A comparison of declarative symbol and gesture use in two children, two bonobos, and a chimpanzee. Lang Commun. 2011;31(1):63–74.

65. Subiaul F, Renner E. Example trials of the cognitive and spatial touchscreen tasks. figshare. Figure. 2020. Available from: https://doi.org/10.6084/m9.figshare.11672304.v1
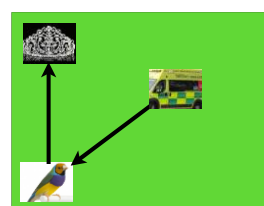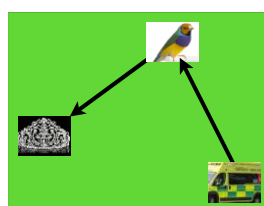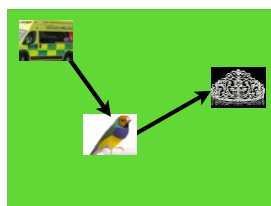
Figure captions

Figure 1. Examples of three trials each of the cognitive task (A) and the spatial task (B). Reproduced from (65).

Figure 2. Summary of child and orangutan performance in the cognitive task by condition. Asterisks inside of the bars denote a condition in which performance was better than Baseline for that group; horizontal lines above bars indicate which conditions are significantly different from each other. The *y* axis shows the first variable in the CI measure (mean correct presses). For **all participants**, this statistic was for **up to the first four** trials in a condition. Error bars indicate standard errors.

Figure 3. Summary of child and orangutan performance in the spatial task by condition. Asterisks inside of the bars denote a condition in which performance was better than Baseline for that group; horizontal lines above the bars indicate which conditions are significantly different from each other. The *y* axis shows the first variable in the CI measure (mean correct presses). For **all participants**, this statistic was for **up to the first four** trials in a condition. Error bars indicate standard errors.

(A)

T1   T2   T*n*

(B)

Cognitive task

Spatial task