Accepted refereed manuscript of: Wilks CEH, Rafetseder E, Renner E, Atkinson M & Caldwell CA (2021) Cognitive Prerequisites for Cumulative Culture are Context-Dependent: Children's Potential for Ratcheting Depends on Cue Longevity. *Journal of Experimental Child Psychology*, 204, Art. No.: 105031. https://doi.org/10.1016/j.jecp.2020.105031

Cognitive Prerequisites for Cumulative Culture are Context-Dependent: Children's Potential for Ratcheting Depends on Cue Longevity

Charlotte E. H. Wilks; Eva Rafetseder; Elizabeth Renner; Mark Atkinson; Christine A. Caldwell

Department of Psychology, Faculty of Natural Sciences, University of Stirling, Stirling, UK

Author Note

Charlotte E. H. Wilks https://orcid.org/0000-0003-0817-6687

Eva Rafetseder https://orcid.org/0000-0002-9816-8607

Elizabeth Renner https://orcid.org/0000-0002-3363-4347

Mark Atkinson https://orcid.org/0000-0002-4242-5459

Christine A. Caldwell https://orcid.org/0000-0001-7362-2554

Elizabeth Renner is now at the Department of Psychology, Durham University

Mark Atkinson is now at the School of Management, University of St Andrews

We have no conflicts of interest to disclose.

Correspondance concerning this article should be addressed to: Charlotte Wilks, Department of Psychology, Faculty of Natural Sciences, University of Stirling, Stirling, UK.

Email: c.e.wilks1@stir.ac.uk

Abstract

Human cumulative culture has been suggested to depend on human-unique cognitive mechanisms, explaining its apparent absence in other species. We show that the potential for exhibiting cumulative culture depends on the cognitive abilities of the agents and the demands associated with using information generated by others' activity. 154 children aged 3-6 years played a searching game ("Find the Treasure"), taking their turn after a puppet demonstrator. The puppet's attempt revealed information about the contents of the locations searched, which could be exploited to target rewarded locations, and avoid unrewarded ones. Two conditions were presented, intended to capture realistic variation in the transience of the cues generated by another individual's activity. In one condition, the puppet's demonstration provided transient information – boxes were opened, seen to be rewarded or not, and then closed. In the other condition the puppet's chosen boxes remained partially open, providing an enduring visible cue as to whether that location was rewarded. Children undertook three trials of varying demonstration success, and we used patterns of performance to infer the potential for improvement over multiple generations of transmission. In the Enduring Cues condition, children's performance demonstrated the potential for cumulative culture. In contrast, in the Transient Information condition, only older children showed improved performances following higher success demonstrations and overall performance was not compatible with the possibility of improvements over generations of social transmission. We conclude that under certain conditions cumulative culture could occur in many species, but in a broader range of contexts in humans.

Keywords: Cumulative culture; cultural evolution; ratchet effect; cognitive development; comparative psychology; social learning

1.0 Introduction

Cumulative culture typically refers to a particular subcategory of cultural evolution, characterised by a directional pattern of change that results in “improvements” (Tennie et al., 2009) or increasingly “preferred” traits (Caldwell, 2018). However, a more precise definition which satisfies the diverse range of scientists interested in this phenomenon is yet to be agreed upon. Mesoudi and Thornton’s (2018) recent sampling of published work outlined four core criteria for cumulative culture: i) a change to a behaviour or cultural product, ii) social transmission of the modified trait, iii) improvement in performance as a consequence of the modification, and iv) iteration of these steps resulting in ongoing improvement over time. These criteria were common across the entire sample and thus can be considered a consensus definition. Yet, they do not encompass additional requirements adhered to by some researchers - classified by Mesoudi and Thornton (2018) as “extended criteria”. For the purpose of adhering to a precise definition, and one which is most representative of that used in the field, we therefore refer to the aforementioned core criteria when using the term cumulative culture. It should however be noted that some other definitions of the concept are more restrictive (e.g. Enquist et al., 2011; Hunt & Gray, 2003; Pradhan et al., 2012; Reindl et al., 2020). The notion of constant improvement has also led to cumulative culture being referred to as the *ratchet effect* (Tennie et al., 2009; Tomasello, 1990). We use these terms interchangeably.

In contrast to its ubiquity in human populations, evidence of cumulative culture in nonhumans is strikingly scarce (e.g. Dean et al., 2014), with some authors proposing that it is unique to humans (e.g. Tomasello et al., 1993). Some have even argued that cumulative culture depends on cognitive mechanisms which themselves are proposed to be unique to humans (Dean et al., 2012; Tennie et al., 2009). However, recent nonhuman evidence from both experimental (Sasaki & Biro, 2017) and field research (Jesmer et al., 2018) appears consistent with Mesoudi and Thornton's (2018) core criteria. These findings in pigeons (Sasaki & Biro, 2017), and bighorn sheep and moose (Jesmer et al., 2018), therefore suggest that cumulative culture is not precluded in nonhumans but may simply be far more restricted in its expression compared with human examples.

In line with this view, experiments on cumulative culture in humans suggest that transmission requirements depend on the context and behaviour in question. For example, Caldwell and Millen (2009), and subsequently Zwirner and Thornton (2015), found that imitation (in the sense of bodily action copying) and teaching are not always required for improvements in performance over multiple transmission events in simple building tasks; exposure to completed products was sufficient. However, imitation may be required for ratcheting during more "cognitively opaque" tasks (e.g. Wasielewski, 2014). There is also little dissent regarding the value of teaching (Morgan et al., 2015) for transmission of skills that are more complex and/or cognitively opaque (Csibra & Gergely, 2009). Consistent with this, Caldwell et al. (2018) found that teaching facilitated the transmission of effective knot-tying techniques for complex knots, whereas simple knots were transmitted equally effectively from exposure to end products alone. Furthermore, Osiurak et al. (2020) showed that a teacher's theory-of-mind ability predicted cumulative performance, but only when the teacher did not have visible access to a learner's actions. Work focussed on the constraints which may underly the tool innovation abilities of individual children, rather than the cumulative result of a social transmission process, is also relevant here; Neldner et al. (2017) provide evidence that the affordance visibility of a tool (whether one can easily perceive how the tool can be used to achieve a desired action, e.g. presence of a visible hook for hooking) may affect the likelihood that task-relevant innovations to the tool occur. The examples outlined above thus raise the possibility that, for a population sharing a common pattern of cognitive abilities and behavioural proclivities, transmission patterns consistent with cumulative culture may be possible for some behaviours but not others. In populations exhibiting a different suite of traits, the range of contexts in which cumulative culture might be manifested could be more extensive, or more limited.

We consider human children to be a particularly interesting group within which these ideas can be tested. Firstly, we know that human adults are capable of cumulative culture. It therefore stands to reason that at some point this becomes possible during human development. However, following the above logic, it is unlikely that we would find evidence of improvements with transmission in extremely young children whose ability to use social information is still developing (e.g. see the age-related improvements in performance documented by Atkinson et al., 2020, for even a trivially simple binomial discrimination task), at least within the context of the kinds of tasks typically presented to adult humans. Nevertheless, it is a reasonable simplifying assumption to consider that children's capabilities typically increase with age, as a consequence of brain maturation and experience. Therefore, we would expect that the range of contexts within which cumulative culture might be manifested should become broader with increasing age. Based on what has been established to date in attempts to demonstrate cumulative culture in children, there is little that would appear to contradict this view, and some tentative evidence in support of it.

Three studies have attempted to test for a ratchet effect in children in ways that could potentially satisfy Mesoudi and Thornton's (2018) core criteria. Reindl and Tennie (2018) adapted Caldwell and Millen's (2008) tower-building task (previously used to demonstrate cumulative culture in adults) for use with 4-5-year-old children. The participants' goal was to build as tall a tower as possible from sticks and plasticine. In transmission chains children were shown the tower built by the previous child or a replica of this. In contrast to Caldwell and Millen's (2008) findings, there was no evidence of a ratchet effect, as tower height did not increase over generations within the chains. Nonetheless, children of this age appear to successively improve on others' solutions under different circumstances. Tennie et al. (2014) presented a task in which children were required to transport rice from one location to another, using any of a set of tools provided. Chains of 4-year-olds, who were able to observe the previous child's attempt before embarking on their own, were not able to improve on the efficiency of solutions that the children in the first generation (i.e. those with no exposure to social information) spontaneously adopted (Tennie et al., 2014). However, when the experimenter acted as the first member of the chain, and introduced an unnecessarily inefficient method, the children in some chains adopted more effective alternatives which were subsequently faithfully transmitted, leading to children in later generations transporting more rice. Similarly, Flynn (2008) found that chains of 2- and 3-year-old children improved the efficiency of tool-use behaviours, eliminating redundant elements that had been incorporated into the original demonstration by the experimenter. Both Tennie et al. (2014) and Flynn (2008) thus show only that children can improve upon abnormally suboptimal methods, as opposed to accumulating increasingly effective solutions - a process termed the “subtractive ratchet effect” (Tennie et al., 2014). Although this behaviour differs from that considered ratcheting by most researchers, these findings suggest that under certain circumstances ratcheting may be possible even in very young children. Creativity and/or insightful innovation, or technical reasoning skills (De Oliveira et al., 2019; Osiurak et al., 2020), may have been required for Reindl and Tennie's (2018) tower-building task. However, in other tasks improvements might readily arise as a consequence of copying error and/or random exploration. In such cases young children may derive benefits from the accumulation of task experience over multiple generations, relative to a baseline of no exposure to social information.

Other studies have tested children in groups, without generational turnover (Dean et al., 2012; McGuigan et al., 2017), claiming that beneficial innovations spread as a consequence of cumulative culture. However, in these studies it is possible that solutions became more effective due to children’s increasing (direct) experience in the test situation: the benefits might not extend to situations involving generational turnover. This caveat aside, the findings are consistent with the idea that ratcheting may be possible from three years onwards, given a task for which the necessary innovations are relatively intuitive. The existing literature therefore suggests that the nature of the task may determine whether social transmission can lead to improvements in performance over multiple generations.

Considering the simplifying assumption that children's capabilities tend to increase with age, it follows that different tasks may have their own age threshold, at which evidence for cumulative culture can be identified.

To date, no study has directly compared age groups in relation to the cultural accumulation of task solutions. This omission is likely to be a consequence of pragmatic considerations, as running such a study using transmission chain or microsociety designs would be extremely challenging from a logistic perspective (Caldwell et al., 2020). Even when not comparing different populations, these methods require large numbers of participants. Each chain or microsociety (consisting of at least three individuals, but usually between five and ten) effectively represents a single sample unit. This therefore necessitates multiple chain/microsociety replicates in order to achieve statistical reliability, and to smooth out the effects of outlier performance by individual participants, which can disrupt an entire chain. Considering the challenges associated with recruitment of a narrowly-defined population (e.g. specific age bands), and the fact that these age bandings would likely need to be even more narrowly defined in order to generate a relatively clean comparison between groupings, it is little wonder few researchers have risen to the challenge.

In the current study we adopt an alternative approach to evaluating the potential for cumulative culture within specified populations, following logic outlined by Caldwell et al. (2020). Rather than relying on sample units consisting of multiple individuals (such as chains, replacement microsocieties, or closed group designs), and considering the effect of generation, or time, on the resulting performance, we test the potential for cumulative culture at the individual level, and consider the effect of exposure to task solutions of varying levels of success. The resulting pattern of performance can be used to infer whether, over multiple transmission episodes, this performance becomes more successful (if, for example – in the case of the most unambiguous evidence – participants were consistently outperforming the demonstration at all levels of task success).

Such an approach would not be possible using the kinds of tasks that researchers have, to date, used to study cumulative culture under laboratory conditions. In most of the experimental literature (and all developmental studies cited above), the method involves one single – usually quite complex – physical puzzle or problem. In these tasks, solutions can vary along a continuum of success, making it possible to track whether performance improves over generations of social transmission. However, using such an approach it is not possible to remove the influence of carry-over effects between multiple attempts and resulting from multiple demonstration exposures. Therefore, due to the inevitable order effects, it becomes impossible to determine how a particular individual would perform in the absence of any other task experience, following exposure to demonstrations of varying levels of success.

Our approach uses a much simpler task (Caldwell et al., 2020) and relies on the fact that the information is strictly episodic, with demonstrations providing information that is specific to the subsequent test trial for that particular problem. Across different problems however, solutions are completely independent from one another. In addition, the nature of our task makes it possible to have precise control over the relative success of the demonstration. This allows us to present a single participant with multiple trials of the same task, without contamination of later trials as a consequence of prior exposure, and with the demonstrations potentially varying along the complete range of possible score outcomes. Thus, from testing at the individual level, we can infer the theoretical outcome of a linear transmission chain of similar performances, in a manner that would not be possible using established methods. For further details about this approach, see Caldwell et al. (2020).

We tested children using a stimulus selection task, a searching game involving an array of stimuli of varying reward value (in this case, simply rewarded or unrewarded), with the objective of maximising reward score. At baseline (i.e. without any social information) it is a guessing game, as stimulus features do not reveal their reward status. However, social information can be provided about reward value. This may require memorising the locations searched by a demonstrator, and whether or not these were rewarded. Alternatively, a demonstrator's activity may leave searched locations marked in some way, with this cue potentially also revealing information about reward value. Whilst the reward value of all stimuli is held constant between demonstration and the participant's selections for a given array (such that information acquired about the reward value of a stimulus during the demonstration can be assumed to hold true for the participant's attempt), each array has its own reward distribution. Therefore, as noted previously, it is possible to expose individual participants to multiple demonstrations of varying success (using different arrays) in order to infer the likely outcome of social transmission through a chain or populations of learners who behave like that particular individual. In addition to looking at children's reward score on the array as a whole, the stimulus selection task enables one to look at how a learner responded to social information regarding the individual rewarded and unrewarded locations selected within the array. This is informative because, if the social information on an array is to be fully utilised and offer the greatest performance benefit, a learner should respond very differently to these two types of information. Looking at this aspect of information use is a further advantage of the stimulus selection task compared to alternative methods in which it can be difficult to assess whether changes in behaviour occur primarily as a consequence of imperfect replication, or active deviation from apparently ineffective elements of a demonstration.

The abstract nature of the stimulus selection task also allows us to manipulate task variables in precise and systematic ways. This is particularly valuable given the likely relationship between task context and the cognitive and behavioural traits required to generate a ratchet effect. In this study we manipulated the longevity of the cues provided by a social demonstration, creating very different demands in terms of memory load and environmental affordance. Cues either remained present in the environment or were present for a limited period - referred to as Enduring Cues or Transient Information conditions, respectively. The distinction between transient and enduring cues corresponds to differences in the information available as a consequence of others' activity in real world social information use. A conspecific's activity might, for example, expose a food source that was previously concealed (e.g. opening of milk bottles by birds (Fisher & Hinde, 1949)), or generate partially processed food items that render the contents more accessible to an inexperienced individual (e.g. pine cone foraging in rats (Terkel, 1996)). Potential foraging locations can also become marked as having been visited simply as a result of visitors leaving perceptible cues at exploited sites (e.g. honeybees and bumblebees leaving scent traces on flowers (Leadbeater & Chittka, 2007)). In these examples, the information endures within the environment which removes both the need to witness the activity, and the potential cognitive burden of storing the information in either short-term memory or working memory, as would be required for transient information. Short-term memory has been defined as the storage of a limited amount of information (Cowan, 2008; Diamond, 2013), and working memory refers to both the storage and processing/manipulation of limited information (Best & Miller, 2010; Cowan, 2008; Diamond, 2013; Garon et al., 2008).

Contrast these examples with typical experimental paradigms used to study social learning (i.e. learning influenced by the observation of, interaction with or behaviour/actions of another animal (Heyes, 1994; 2012)) in nonhuman primates and human children (e.g. see Whiten et al. (2009) for a review). These studies typically involve a single apparatus, operated by only one individual at a time, and the behaviour of interest is usually a specific method or action which can be performed on the apparatus. The social information available from others' interaction with the apparatus is therefore available only relatively instantaneously and leaves no lasting trace. Furthermore, such information requires storage since simultaneous activity is not possible.

The behavioural and cognitive requirements of social information use thus depend critically on contextual details such as those outlined above. Our aim was therefore to investigate the potential for ratcheting in children of a range of ages, under conditions of varying longevity of social information cues. Our expectation was that in certain populations (i.e. particular age groups, in this instance), a ratchet effect might be possible under conditions involving enduring environmental cues, whilst being precluded when cues were transient.

We presented the searching task to children by framing it as a challenge to find hidden treasure, a context we believed would be understandable and motivating for them. The reward ("treasure") was always hidden in one of three locations (of the same colour, Figure 1a), and a total of three rewards could be searched for on each of three different, nine-chest arrays (Figure1b). This meant that the total score on each array could vary between 0 (no rewards found) and 3 (all rewards found). This allowed us to create demonstration trials of varying success, and to study how well children made use of social information about locations of rewards (arising from correct choices) and locations to be avoided (arising from incorrect choices). As well as predicting effects of age and cue longevity on overall task performance, we intended to use children's scores on the different array types to infer the likely outcome of a series of social transmission events and assess the potential for ratcheting (in line with the logic in Caldwell et al., 2020). Children were therefore exposed to demonstrations reflecting three different success levels. In order to be classified as displaying the potential for ratcheting on this task, an individual (or in this case, age group) would need to perform significantly above chance level overall, showing some ability to benefit from the social information. However, the success of the individual or group in question would also need to be related to the success of the demonstration, such that higher scores followed higher scoring demonstrations. If demonstration success has no bearing on the child's score then clearly no ratcheting would be possible, since social learners derive no additional value from exposure to beneficial modifications, resulting in a performance plateau. Indeed, assuming linear transmission, demonstrating a potential for ratcheting requires participant performance to be above the level of the demonstration for a minimum of two successive simulated transmission events. This criterion ensures that the evidence is consistent with not just benefits from exposure to social information, but also the accumulation of benefits over successive transmission episodes. In the current study therefore, our criteria for potential for ratcheting were outperformance of both a chance-level (1-rewarded) demonstration, equivalent to the typical outcome of naïve exploration, *and* an above-chance demonstration (2-rewarded).

We predicted that children would perform better in the Enduring Cues condition, compared with the Transient Information condition. As previously stated, using transient information is expected to place greater demands on memory. In order to achieve the best scores in our task we suspect that working memory may be utilised because both storage and manipulation of information are likely to be required (Best & Miller, 2010; Cowan, 2008; Diamond, 2013; Garon et al., 2008). In order to make use of information from unsuccessful demonstration attempts (i.e. the revealing of unrewarded locations) one would presumably need to hold a memory of the unrewarded location in mind whilst deciding which of the remaining available locations to search, a similar process to that required in complex working memory tasks (Garon et al., 2008). We thus predicted that children would perform better with increasing age (in line with developments in working memory (Best & Miller, 2010; Diamond, 2013, Garon et al., 2008)), across both conditions. Additionally, we used the criteria described above to estimate the extent to which age groups might contribute to the accumulation and retention of beneficial modifications through social transmission, under both conditions.

2.0 Materials and Methods

2.1 Participants

We collected data from 163 children aged three to seven across the two conditions. The one seven-year-old was placed into the same category as six-year-olds for descriptive statistics and t-tests but all further analysis used age in days. Eighty-three children were recruited from Glasgow Science Centre and a further eighty from a primary school in Bradford, UK (further details in SI). Nine children were excluded for the following reasons: missing date of birth from the consent form (age 3, female), failure to fully comply with task instructions (n=4; aged 3, 5 and two aged 4; male), inability to understand the task due to language difficulty (age 3, male) and experimenter error (n=3, aged 6 and two aged 4, female). The final sample consisted of 154 children aged three to seven (M= 59 months, range= 38 – 84, SD= 13, 88 female). There were 76 and 78 children in the Enduring Cues and Transient Information conditions, respectively.

Ethics statement: This research was approved by the University of Stirling, General University Ethics Panel (references: GUEP40 & GUEP289). Written, informed consent was obtained from the parent or guardian of all children prior to their participation. Children were asked if they would like to participate, were continuously monitored for assent and were rewarded with a sticker regardless of task completion.

2.2 Apparatus

A large parrot hand puppet (The Puppet Company©) was used as a demonstrator, performing all social selections. The choice stimuli were wooden treasure-chests (82mm, 52mm, 46mm), containing either 30 (30mm x 30mm) squares of felt treasure (coloured to match the chests) as reward stimuli, or scrunched-up newspaper as a cue to the absence of a reward. Three chests of each of the following colours: red, yellow, blue, green, purple, orange, white, brown, and pink were presented in groups of three, nine-chest arrays (Figure

1a). Each array was placed in turn onto a laminated treasure map measuring 426mm x 482mm (Figure S1). Two boards, each containing nine coloured squares (corresponding to the nine different colours of treasure to be found), were used to keep score (Figure S2). Each time treasure was found by the puppet/child it was stuck to the corresponding colour on their score board.

2.3 Procedure

In the school, testing took place in a quiet room adjacent to the classroom. At Glasgow Science Centre, testing was carried out in a public space, separated from the main museum space. In the museum only, less confident children were accompanied in the testing area by a parent or guardian, who was instructed not to provide the child with any task-relevant assistance. In both locations the experimental task was carried out on a table-top with the child seated next to the experimenter and opposite the puppet (operated by a research assistant), Figure S2. A verbal script was used by the experimenter (see SI).

Children were asked if they would like to play a game in which the goal was to try to find more treasure than the puppet. A series of nine selections were made over three stimulus arrays (three selections per array, Figure 1b) and each selection was made from a choice of three chests of the same colour, which could be either rewarded (contain treasure) or unrewarded (not contain treasure). An array could therefore be viewed as a three-choice search task consisting of three puppet selections, and three child selections. In the Transient Information condition only, each child selection was preceded by two memory questions to assess memory of the puppet’s selections - note that the puppet performed all three of its selections before the child began their turn. The child’s turn then consisted of the following steps: memory questions for the puppet’s first selection, child’s first selection, memory questions for the puppet’s second selection, child’s second selection etc.

Each child was assigned to either a Transient Information or an Enduring Cues condition. In the Transient Information condition treasure chests were fully closed following the puppet's selection, therefore information on the contents was only available for the duration of the choice itself. In comparison, chests in the Enduring Cues condition were left partially open so that the contents remained on view during the puppet's remaining selections on that array, and throughout the child's selections. The condition to which children were assigned was determined according to a combination of factors, including the maintenance of balanced numbers across conditions within age groups.

Children were first asked to look inside the three chests of the first colour group in the current array (e.g. red in array 1 – Figures 1a and S1) before placing them into a sack. They were then asked to do the same with the chests of the second and then third colour groups. This was to ensure children understood that only one chest of each colour contained the treasure, and that the remaining two contained newspaper. The nine chests (of the three different colour groups in the current array) were then mixed up inside the sack so that the child could not know which chests contained the treasure. Next, the child was asked to remove the three chests of the first colour group (e.g. three red chests for array 1) and place these on large X's which were marked in a line at the top (left, centre and right) of a treasure map (X's for the remaining two colour groups were in lines across the centre and bottom of the map) ready for the puppet to make its first selection, which it then did. Following this selection, chests of the second colour group were placed onto the map and so on until the puppet had made all three choices.

2.3.1 Puppet Selections

Each array consisted of three consecutive puppet selections which were made before the child made any selections (see section 2.3.3). The assistant controlling the puppet knew which chests were rewarded due to discreet pencil marks drawn on the back of the chests. This enabled them to select a rewarded or unrewarded chest as predetermined. Prior to a puppet's selection, children were told that "Polly the Parrot" was going to look for the treasure (of one particular colour e.g. red for selection 1, array 1) and the puppet then selected one chest by touching it with its beak; the child was prompted to open this chest. If the puppet was successful, the child was asked to take a piece of treasure and give it to the puppet, leaving the chest full of the remaining pieces of treasure. The puppet then placed its piece of treasure onto the relevant coloured square on its score board. If the puppet was unsuccessful nothing was removed from the chest. In the Transient Information condition chests were closed immediately following the puppet's selection, and in the Enduring Cues condition they were left partially open so that the rewarded/unrewarded contents were clearly visible.

2.3.2 Memory Questions

On each array, the three consecutive puppet selections were followed by three consecutive child selections. Each child selection was preceded by two memory questions (Figure 1b) - "Where did Polly the Parrot look for the [*red*] treasure?" and – "Did Polly find the [*red*] treasure?" Only their answer (given either verbally or by pointing) to the first question was used in the analysis because it was possible that children were using Polly's score board to determine whether she had found the treasure. These questions were asked in the Transient Information condition only and all children responded (see SI for the results).

2.3.3 Child Selections

Immediately following each set of memory questions children were asked to search for treasure in chests of the corresponding colour group. The procedure was as follows for each array: child memory questions 1, child selection 1, child memory questions 2, child selection 2, child memory questions 3 and child selection 3 (Figure 1b). Child selection involved a child choosing and opening one of the three chests in a colour group, as the puppet had done. If the child was successful, they were prompted to take a piece of treasure and place it onto their score board before the chest was closed. If they were unsuccessful nothing was removed from the chest and it was closed.

The total number of rewarded social selections in each of the three arrays differed. The puppet found one rewarded chest (in total across the three colour groups in an array) on one array, two rewarded chests on another, and three rewarded on a further array and thus scored 1/3, 2/3 and 3/3 - a total score of 6/9 which corresponded to six pieces of treasure found. The order of these three scores across the three arrays was counterbalanced: each of the six possible orders was assigned to participants within each age category in a randomised order. On the arrays in which 1/3 or 2/3 of the colour groups were rewarded we counterbalanced the position of the rewarded colour group(s) (1^{st}, 2^{nd} or 3^{rd}) and assigned all possible combinations to participants within each age category in a randomised order.

The following responses were live coded: the puppet's selections (necessary because, although selection of a rewarded/unrewarded chest was predetermined, the exact unrewarded chest chosen was not), the child's responses to the two memory questions (although only answers to the first were used in the analysis, see SI), and the child's selections. Following a rewarded information trial, the optimal response was to repeat the puppet's selection and hence also find the treasure. Following an unrewarded information trial, the optimal response was not to repeat the puppet's response and to select one of the two alternative chests of the same colour, which resulted in a 50% chance of finding the treasure. If children always repeated the puppet's responses following rewarded trials, but shifted to different chests following unrewarded trials, we would expect an average child success score of 7.5. This would reflect six points for selecting the same chests as the puppet for the rewarded trials, plus an average score of 1.5 out of 3 for the unrewarded trials (50% correct).

3.0 Results

We were interested in the extent to which children used the social information provided by the puppet in our Transient Information and Enduring Cues conditions. We therefore measured information use with a number of dependent variables (discussed below). We were also interested in children's memory for the locations selected by the puppet in the Transient Information condition (see SI). *p*-values < .05 were taken as statistically significant across all analyses. All generalised linear mixed effects models (GLMM) and generalised linear models (GLM) were carried out with either the logit link (binomial data, family = binomial) or log link (count data, family = poisson) and the lme4 package (lmer and glm functions for GLMMs and GLMs, respectively) using R (R Core Team, 2018). Our default choice for the random effects structure for each model included by-participant random slopes for variables which varied within participant, following Barr et al. (2013), to keep random effects structures "maximal" where possible. Where the "maximal" model resulted in non-convergent or singular fit models, random slopes were removed followed by random intercepts where necessary until a convergent, non-singular model was obtained. We list the final structure for each model, including all random intercepts and slopes, under the relevant sections below and in the SI (Table S3). The following independent variables were used across our "Optimal Response Count" and "Repeating" analyses (below):

Between subjects - age (thousands of days, centred), condition (sum coded: Transient Information -1, Enduring Cues 1).

Within subjects - information type (sum coded, unrewarded set to -1, rewarded to 1).

3.1 Optimal Response Count

In this analysis, children were awarded points for repeating rewarded selections, and not repeating unrewarded selections (see SI for further justification). As there were nine puppet selections, the maximum possible optimal response count was nine (Figure 2: mean counts by age and condition). Note that this count differed from a child's success score on the game because, rather than simply being a sum of the number of pieces of treasure found, this accounted for the fact that it was possible to correctly not copy an unrewarded puppet selection and yet fail to find the treasure. However, despite giving a broad overview of information use, the optimal response count was a rather crude measure because it did not distinguish between what children did following rewarded and unrewarded social information trials, or following 1/3, 2/3 and 3/3 rewarded arrays.

We performed a generalised linear model with optimal response count as the dependent variable and age, condition, and the interaction of age and condition as fixed effects. There were main effects of age ($p = 0.008$) and condition ($p < .001$): optimal response count was higher in older children and in the Enduring Cues condition (Figure 2). The interaction between age and condition was approaching significance (p = 0.063).

Despite the non-significant interaction effect between age and condition in our previous model, we further examined the effect of age on optimal response count in our two conditions. These analyses were considered worthwhile due to the relevance to our hypotheses of different age effects across the two task conditions. However, these models should be regarded as purely exploratory, and interpreted with caution accordingly. We again used a dependent variable of optimal response count, but condition was removed from the fixed effects and we split the data up by condition prior to running the models. These models therefore had age only as a fixed effect but were otherwise identical to the above. In the Enduring Cues condition, there was no main effect of age on optimal response count ($p = 0.539$). Contrastingly, there was a highly significant main effect of age in the Transient Information condition ($p = 0.003$): higher optimal response count in older children (Figure 2).

3.2 Repeating

Repeating the puppet's rewarded selections, and not repeating the puppet's unrewarded selections, are both correct strategies but pose different demands. While the former requires one to remember a rewarded selection and repeat it, the latter requires one to remember an unrewarded selection and avoid repeating it. If the social information was being used effectively, we would expect high and low levels of repeating following rewarded and unrewarded puppet selections, respectively. Repeating was binary coded, “1” for a repeat and “0” for no repeat, for all nine responses made by each child (see Table 1 for mean proportion of responses repeated by information type, age and condition and SI for further details on this measure).

3.2.1 Effects of Information Type, Age, and Condition

We performed a GLMM with repetition of responses as the (binary) dependent variable; and information type, age, condition, and their interaction, as fixed effects. We included a by-participant random slope for information type and random intercepts for participant, reward position (reward located on the left, centre or right of the row of chests) and trial number (1-9). There were main effects of information type ($p < .001$) and age ($p = 0.027$): more repeating following rewarded trials and in older children. There was no main effect of condition ($p = 0.098$) but there was a significant interaction between information type and condition ($p < .001$), showing that information type affected the likelihood of a response being repeated more in the enduring condition than the transient. There was no significant interaction effect between information type and age ($p = 0.051$), although this fell just short of statistical significance. There was no interaction between condition and age ($p = 0.648$) or information type, condition and age ($p = 0.263$).

3.2.2 Effects of Information Type and Age in the Enduring Cues and Transient Information Conditions

We ran two further models, with a view to understanding the differing effects of information type on repeating performance within each condition. These models had the dependent variable of repetition of responses (as above), but we removed condition from the fixed effects and instead split the data by condition. This left information type and age only as fixed effects. The model for the Enduring Cues condition had a by-participant random slope for information type and random intercepts for participant, reward position and trial number, as in the previous model. The model for the Transient Information condition had random intercepts for trial number and participant. We also added a random intercept for "remembered" (0 or 1 as per the child's response to the first memory question) in order to account for any effect of ability to remember which chest the puppet selected on repeating. For the Enduring Cues we found a main effect of information type ($p < .001$): more repeating following rewarded trials than following unrewarded trials (Figure 3a). However, there was no evidence for an effect of age ($p = 0.400$), or an interaction between age and information type ($p = 0.768$). In the Transient Information condition (Figure 3b) there was also a main effect of information type ($p < .001$) and no main effect of age ($p = 0.470$). However, there was a significant interaction between age and information type ($p < .001$), which indicated that the effect of information type on repeating was stronger in the older children, compared with the younger. There was also a large variance (2.045) associated with the remembered variable.

3.2.3 Effects of Age and Condition by Information Type

We ran two further models, with a view to understanding the differing effects of age on repeating performance for rewarded and unrewarded information. We split the data by information type. These models therefore still had a dependent variable of repetition of responses but fixed effects of age and condition only. The model for the rewarded information had a random intercept for trial number and that for unrewarded had no random effects. For the rewarded selections we found main effects of age ($p < .001$) and condition ($p < .001$) - more repeats in older children, and in the Enduring Cues condition, but no interaction effects ($p = 0.431$). For unrewarded selections there was a main effect of condition ($p < .001$): fewer repeats in the Enduring Cues condition. However, there was no main effect of age ($p = 0.885$) or interaction of age and condition ($p = 0.377$). These models therefore showed that there was increased repeating with age for rewarded selections but not decreased repeating with age for unrewarded selections despite both repeating after rewarded and not repeating after unrewarded being effective behaviours (we return to this point in our analysis of the second memory question, see SI). Overall repetition of responses (more repeats after rewarded selections and fewer repeats after unrewarded selections) was more accurate in the Enduring Cues condition (Figure 3a) in which the information was more freely assessable.

3.3 Potential for Ratcheting (PFR[1])

We grouped children according to chronological age, using one-year bandings, and classified the performance of each group, in each condition, according to the PFR continuum described below. For an age group to be defined as showing PFR in a linear transmission chain the level 3 criteria needed to be fulfilled.

Our PFR classification continuum was as follows: Level 0: Chance-level performance across all trials, regardless of demonstration success (no benefit from social information); Level 1: Above chance performance overall, but scores no higher, on average, following higher success demonstrations (2/3 and 3/3), compared with lower (1/3) (social learning benefit unlinked to demonstration success); Level 2: Above chance performance overall, with scores higher, on average, following higher success demonstrations (2/3 and 3/3), compared with lower (1/3), without outperforming higher success demonstrations (2/3) (social learning benefit linked to demonstration success); Level 3: Outperform demonstrations by scoring higher following both chance-level (1/3) and higher success (2/3) demonstrations (potentially supporting ratcheting under linear transmission).

For level 3 ratcheting, we included the first of these criteria because outperforming a chance-level (1/3) demonstration shows that the ratchet effect can get off the ground, with naïve exploration as a starting point. The second criterion verifies that subsequent transmission events can result in further performance improvements, by evaluating whether social learners also outperform a demonstration equivalent to expected optimal performance resulting from exposure to naïve exploration (2/3). This pattern of performance would

[1] We hereafter use the abbreviation “PFR” to denote “Potential for Ratcheting”.

support ratcheting even under linear (single cultural parent) transmission. However, under some circumstances, ratcheting may occur without outperformance of an individual model's higher success demonstrations. For example, in a situation in which there is more than one cultural model, performance may improve following the use of multiple pieces of valuable information from different models. Such non-linear transmission scenarios are captured by our level 2 criteria. Under these level 2 criteria, it had to be clear that children were performing better following high-success demonstrations (2/3 and 3/3), compared with lower-success demonstrations (1/3), illustrating that success was linked to the quality of information available. Although children achieving this level were not able to outperform the higher-level 2/3 demonstration directly, the benefit they got from higher success demonstrations means that, at a population level, performance could potentially improve over generations if learners were exposed to multiple demonstrations.

In order to classify children of each age group according to our PFR continuum we gave children three success scores, each representing their aggregated performance on one of the different success level arrays (1/3, 2/3 and 3/3 rewarded). Children were given a point each time they found the treasure and zero points if they did not. They could therefore score up to one point on each colour group, a maximum of three points on each array and nine points over the entire task. We calculated the mean success scores for each age group and condition for the entire task (success score/9); the one-, two-, and three-rewarded arrays; and the mean of the two and three rewarded (Table 2 and Figure 4). We then performed one-tailed *t* tests and used these results to classify the performance of each group according to our PFR continuum (Figure 5). We employed a "strict" and a "less strict" method when assessing whether children fulfilled the requirements necessary to be categorised as a particular level. For the "strict" criteria, children from a particular age group needed a mean success score which was significantly higher than the benchmark outlined in a particular level criterion but for the "less strict" criteria a numerical difference, but not significance, was required. We report the less strict criteria in full here, due to the crossing (or otherwise) of the threshold being perhaps more relevant for this purpose than the degree of confidence in that conclusion. However, results according to the strict criteria are reported in full in SI, and both are displayed on Figure 5.

According to the less strict criteria, in the Enduring Cues condition, children of all ages were categorised as level three (Figure 5) because the mean success scores for the 1/3 demonstration were greater than chance-level (1) and the mean success scores for the 2/3 demonstration were greater (although not all *significantly* greater) than 2 (Table 2 and Figure 4). However, in the Transient Information condition, although the mean success scores were greater than chance-level for each age group, the mean success scores for the 2/3 demonstrations were not >2 for any age group (Table 2 and Figure 4). Furthermore, mean success scores for the 2/3 and 3/3 demonstration types combined (Table 2) were above 2 at age six only; therefore children aged three, four and five were classified as level one and children aged six as level two in the Transient Information condition (Figure 5).

4.0 Discussion

We used a novel method of testing capacities for cumulative culture, which allowed us to evaluate performance at the individual level. This gave much greater sensitivity for identifying age effects in children and was motivated by our aim of testing our ideas about the constraints on cumulative culture. Specifically, we had reasoned that whether cumulative culture would be manifested within a given population depends on the demands of social information use in the specific learning context in question, as well as the characteristics of members of the population. It followed that, in certain populations, performances indicative of the potential for cumulative culture might be observed in some contexts but not others. In line with this reasoning, we had predicted that children’s age would be positively related to their scores on our proxy measures of potential for ratcheting. We also expected that, in a version of the task with reduced cognitive demands, scores would be higher than they were in an otherwise similar task that imposed significant demands on memory. This was expected to result in patterns of performance indicative of potential for cumulative culture being more prevalent under reduced cognitive demands, and apparent from a younger age.

Our results broadly supported these predictions, in that scores were higher in our Enduring Cues condition relative to the Transient Information. Furthermore, in the latter it was only in the oldest age group (6 years) that we found any suggestion of the potential for a ratchet effect and our optimal response count increased with age in this condition only. When looking at information type separately we found that children's ability to utilise information about rewarded selections increased with age, but not for unrewarded, which did not significantly improve (we discuss how this may relate to the result of the first memory question in the SI). We return to evaluate the rationale and effectiveness of these metrics later in the discussion, and first consider the validity of the contrasting conditions, and the distinction between them.

It is important to note that in our Enduring Cues condition, which left visible evidence of a location's reward value, the "demonstration" element (i.e. observing the puppet's choices) is likely redundant, or influential only inasmuch as it set the information in context. In this condition the demonstration may have functioned to reinforce understanding of the goal of the task and the desirability of the contents of the boxes (coloured pieces of "treasure" versus crumpled newspaper). However, assuming this understanding was in place, the knowledge that the cues arose as a consequence of another's activity was not required in order for the children to make effective use of the information. Arguably therefore, the Enduring Cues condition did not involve "social" learning at all. It is likely that the children would have performed just as well, or possibly better, had the cues been generated by their own individual exploration, rather than the social demonstration. This applies to both conditions.

Other work in our group (Atkinson et al., 2020; Renner et al., submitted) has investigated children's performance following exposure to information acquired from a social demonstration compared with individual exploration, in logically similar – although simplified – tasks. Both found that children's use of information was largely unaffected by source. However, in the current study, we were not concerned with the question of children's understanding of the source of the information, nor whether their responses might be dependent on the nature of that source. Here we simply wished to evaluate their ability to use such information in ways which could potentially lead to ratcheting of task performance based on cultural transmission, and the conditions under which this might be possible. However, as noted in section 1.0, the definitions of cumulative culture used by some researchers are more restrictive than ours, which adhered to Mesoudi and Thornton's (2018) core criteria. Such researchers may include additional requirements (e.g. Mesoudi and Thornton's (2018) extended criteria) and thus not regard our task as suitable for detecting cumulative culture. This is consistent with our view that the cognitive demands associated with cumulative culture may differ depending on the precise context of the behaviour being transmitted - more restrictive definitions of cumulative culture likely involve situations with a different set of cognitive demands.

As outlined in our introduction, there are now examples of information ratcheting in migratory species previously unstudied in relation to cumulative culture – pigeons (Sasaki & Biro, 2017) and bighorn sheep and moose (Jesmer et al., 2018). In both, benefits to subsequent generations are transmitted by virtue of the grouping tendencies of the species in question, and those benefits likely accumulate as a consequence of direct feedback to the individuals based on relatively random deviations from the acquired behaviour. Therefore, there appears to be no necessity for complex cognitive abilities to either extract cloaked information or to innovate new solutions. We consider these examples to be somewhat comparable to our Enduring Cues condition in which children had full access to information which could be utilised to improve their success score. Our results allow us to infer what would occur in a social transmission scenario involving a series of children attempting the task once each in succession. Results from the Enduring Cues condition suggest that children would typically achieve maximum score (finding all rewards) following just a few generations of transmission (probably by the fourth or fifth child to attempt the task). In contrast, results from the Transient Information condition suggest that information about reward locations would not be reliably retained during transmission, preventing accumulation of benefits. If cognitive abilities not within a child's or species' repertoire are needed in order to access certain information (e.g. memory ability in our Transient Information condition) then this places limits on the potential for ratcheting over generations. However, it follows that ratcheting may be perfectly possible when such constraints are absent. Our finding that other aspects of cognition (i.e. not specifically socio-cognitive) may place constraints on the potential for cumulative culture is consistent with emerging literature. For example, Osiurak & Reynaud (2020) have highlighted the importance of technical reasoning skills for cumulative culture of technology. And Tennie and colleagues (e.g. Tennie et al., 2009) have proposed the "Zone of Latent Solutions" hypothesis which posits that limits on nonhumans' individual learning capacities will place constraints on their cultural behaviours.

It has been proposed that enhanced working memory in the human lineage has enabled cumulative culture, however this opinion is based on evidence from the archaeological record rather than measured differences in cognition (Balter, 2010; Coolidge & Wynn, 2001; Wynn & Coolidge, 2011). Comparing human and nonhuman memory directly remains challenging due to the small amount of comparative research available for analysis. It is likely that some nonhuman mammals have a similar memory storage capacity to that of humans, but that humans can represent concepts within memory differently and have a better ability to deploy attention and resist interference (for a systematic analysis of evidence to date see Carruthers, 2013). Humans can also extend their own memory capacity through the development of external storage systems such as writing (which acts to enhance working memory through literacy learning, Wynn & Coolidge, 2011), mnemonic devices (Jurowski et al., 2015) or inner speech which allows for the rehearsal of items held in memory (Carruthers, 2013; Cowan, 2008). Such strategies have been described as “cognitive offloading” and children’s tendency to use them increases with age (Armitage et al., 2020). It may therefore be the case that young children, who are yet to fully develop metacognitive storage strategies or the ability to devise cognitive offloading strategies (Bulley et al., 2020), have a greater reliance on their raw memory storage capacity when solving tasks such as that in this study, and may therefore perform similarly to some nonhumans. If nonhuman primates were to complete a similar task, assuming the relevance of the cues had been learned, we would expect them to show PFR in the Enduring Cues (as children did from age 3) due to the low memory load requirements. Comparatively, if performance on the Transient Cues relies on using strategies such as inner speech, or if humans do have greater memory capacities, it is unlikely that other species would show PFR in this condition.

In this study we used a simple stimulus selection task to minimize incidental task demands that might constrain performance, and to allow us to manipulate key variables of interest. This extreme simplification means that the task probably deviates in significant ways from most real-world cases of social learning. For example, the “decision” as to whether the social information should be repeated is binary: either repeat the demonstrator’s choice or choose an alternative location. Furthermore, the payoffs of the demonstrator’s choices are completely transparent, and the structure of the reward landscape is known (i.e. one of the three options is rewarded), even if the location of the reward is not. Therefore, it is possible to work out the optimal response, subject only to the availability of the information either in the environment or from internal storage. Whilst it could be argued that these simplifications are unrealistic, and that our paradigm reveals little about social learning in the real world, we contend that the demands we eliminated (opaque payoff information, ill-defined solution space, and unknown reward landscape) are not peculiar to social learning. There is therefore no reason that social learning paradigms must encompass these associated challenges. Task simplification allowed us to establish that a ratchet effect was possible, in principle, even in very young children. However, the stimulus selection task also offers scope to investigate alternative questions related to social information use e.g. ritual or strategy use (Kapitány et al., 2018).

4.1 Conclusions

Based on children's PFR performance in our Enduring Cues and Transient Information conditions it appears that superior cognitive capacities (such as working memory) were required to access (and therefore use) the information provided in the Transient Information condition compared to the more accessible information in the Enduring Cues. We conclude that the potential for a ratchet effect (whether considering children of a particular age, or members of a nonhuman species), depends on the type of information itself, as well as the cognitive capacities required to access that information.

Funding

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme. Grant agreement No. 648841 RATCHETCOG ERC-2014-CoG.

Acknowledgements:

Assisted with study design: Gemma Mackintosh

Assisted with data collection: Kirsten Blakey, Kymberly Craig, Tamsyn Pereira, Eilidh Shearer, Mirela Stefanova and Christine Wilks

References

Armitage, K. L., Bulley, A., & Redshaw, J. (2020). Developmental origins of cognitive offloading. *Proceedings. Biological Sciences*, *287*(1928), 20192927. https://doi.org/10.1098/rspb.2019.2927

Atkinson, M., Renner, E., Thompson, B., Mackintosh, G., Xie, D., Su, Y., & Caldwell, C.A. (2020). Robust, source-independent biases in children's use of socially and individually acquired information. *Journal of Experimental Psychology: General*. 10.1037/xge0000959

Balter, M. (2010). Did working memory spark creative culture? *Science*, *328*, 160–163. https://doi.org/10.1126/science.328.5975.160

Barr, D.J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Best, J., & Miller, P. (2010). A Developmental Perspective on Executive Function. *Child Development*, *81*(6), 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Bulley, A., McCarthy, T., Gilbert, S. J., Suddendorf, T., & Redshaw, J. (2020). Children Devise and Selectively Use Tools to Offload Cognition. *Current Biology*, 1–8. https://doi.org/10.1016/j.cub.2020.06.035

Caldwell, C. A. (2018). Using Experimental Research Designs to Explore the Scope of Cumulative Culture in Humans and Other Animals. *Topics in Cognitive Science*, 1–17. https://doi.org/10.1111/tops.12391

Caldwell, C. A., Atkinson, M., Blakey, K. H., Dunstone, J., Kean, D., Mackintosh, G., Renner, E., & Wilks, C. E. H. (2020). Experimental assessment of capacities for cumulative culture: Review and evaluation of methods. *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*, 1–15. https://doi.org/10.1002/wcs.1516

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., Renner, E., & Atkinson, M. (2018). Human Teaching and Cumulative Cultural Evolution. *Review of Philosophy and Psychology*, *9*, 751–770. https://doi.org/10.1007/s13164-017-0346-3

Caldwell, C. A, & Millen, A. E. (2009). Social Learning Mechanisms and Cumulative Cultural Evolution. *Psychological Science*, *20*, 1478–1483. https://doi.org/10.1111/j.1467-9280.2009.02469.x

Carruthers, P. (2013). Evolution of working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 10371–10378. https://doi.org/10.1073/pnas.1301195110

Coolidge, F. L., & Wynn, T. (2001). Executive functions of the frontal lobes and the evolutionary ascendancy of Homo sapiens. *Cambridge Archaeological Journal*, *11*, 255–260. https://doi.org/10.1017/S0959774301000142

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in brain research*, *169*, 323–338. https://doi.org/10.1016/S0079-6123(07)00020-9

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*, 148–153. https://doi.org/10.1016/j.tics.2009.01.005

De Oliveira, E., Reynaud, E., & Osiurak, F. (2019). Roles of Technical Reasoning, Theory of Mind, Creativity, and Fluid Cognition in Cumulative Technological Culture. *Human Nature*, 326–340. https://doi.org/10.1007/s12110-019-09349-1

Dean, L.G., Kendal, R. L., Schapiro, S. J., Thierry, B., & Laland, K. N. (2012). Identification of the social and cognitive processes underlying human cumulative culture. *Science*, *23*, 1114–1118. https://doi.org/10.1126/science.1213969

Dean, L.G., Vale, G. L., Laland, K. N., Flynn, E., & Kendal, R. L. (2014). Human cumulative culture: A comparative perspective. *Biological Reviews*, *89*, 284–301. https://doi.org/10.1111/brv.12053

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Enquist, M., Ghirlanda, S., & Eriksson, K. (2011). Modelling the evolution and diversity of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1563), 412–423. https://doi.org/10.1098/rstb.2010.0132

Flynn, E. (2008). Investigating children as cultural magnets: do young children transmit redundant information along diffusion chains? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*, 3541–3551. https://doi.org/10.1098/rstb.2008.0136

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Heyes, C. (1994). Socal learning in animals: Categories and mechanisms. *Biology Review*, *69*, 207–231. https://doi.org/10.1111/j.1469-185X.1994.tb01506.x

Heyes, C. (2012). Grist and mills: on the cultural origins of cultural learning. *Philosophical Transactions of the Royal Society B*, *367*(1599), 2181–2191. https://doi.org/10.1098/rstb.2012.0120

Hunt, G. R., & Gray, R. D. (2003). Diversification and cumulative evolution in New Caledonian crow tool manufacture. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1517), 867–874. https://doi.org/10.1098/rspb.2002.2302

Jesmer, B. R., Merkle, J. A., Goheen, J. R., Aikens, E. O., Beck, J. L., Courtemanch, A. B., Hurley, M. A., McWhirter, D. E., Miyasaki, H. M., Monteith, K. L., & Kauffman, M. J. (2018). Is ungulate migration culturally transmitted? Evidence of social learning from translocated animals. *Science*, *361*, 1023–1025. https://doi.org/10.1126/science.aat0985

Jurowski, K., Jurowska, A., Krzeczkowska, M., Jurowski, K., Jurowska, A., & Krzeczkowska, M. (2015). Comprehensive review of mnemonic devices and their applications : state of the art. *International E-Journal of Science, Medicine and Education*, *9*, 4–9.

Kapitány, R., Davis, J., Legare, C.H., Nielsen, M. (2018). An Experimental Examination of Object-Directed Ritualized Actions in Children Across Two Cultures. *PLoS ONE*, 13(11), 1-14. 10.1371/journal.pone.0206884

Leadbeater, E., & Chittka, L. (2007). The dynamics of social learning in an insect model, the bumblebee (Bombus terrestris). *Behavioral Ecology and Sociobiology*, *61*, 1789–1796. https://doi.org/10.1007/s00265-007-0412-4

McGuigan, N., Burdett, E., Burgess, V., Dean, L., Lucas, A., Vale, G., & Whiten, A. (2017). Innovation and social transmission in experimental micro-societies: Exploring the scope of cumulative culture in young children. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*. https://doi.org/10.1098/rstb.2016.0425

Mesoudi, A., & Thornton, A. (2018). What is cumulative cultural evolution? *Proceedings of the Royal Society B: Biological Sciences*, *285*. https://doi.org/10.1098/rspb.2018.0712

Morgan, T. J. H., Uomini, N. T., Rendell, L. E., Chouinard-Thuly, L., Street, S. E., Lewis, H. M., Cross, C. P., Evans, C., Kearney, R., De La Torre, I., Whiten, A., & Laland, K. N. (2015). Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature Communications*, *6*, 4–11. https://doi.org/10.1038/ncomms7029

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

Osiurak, F., & Reynaud, E. (2020). The elephant in the room: What matters cognitively in cumulative technological culture. *Behavioral and Brain Sciences, 43*, E156. doi:10.1017/S0140525X19003236

Osiurak, F., De Oliveira, E., Navarro, J., & Reynaud, E. (2020). The castaway island: Distinct roles of theory of mind and technical reasoning in cumulative technological culture. *Journal of Experimental Psychology: General*, *149*(1), 58–66. https://doi.org/10.1037/xge0000614

Pradhan, G. R., Tennie, C., & van Schaik, C. P. (2012). Social organization and the evolution of cumulative technology in apes and hominins. *Journal of Human Evolution*, *63*(1), 180–190. https://doi.org/10.1016/j.jhevol.2012.04.008

R Core Team (2018). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org/.

Reindl, E., & Tennie, C. (2018). Young children fail to generate an additive ratchet effect in an open-ended construction task. *PLoS ONE*, 13(6), e0197828 https://doi.org/10.1371/journal.pone.0197828

Reindl, E., Gwilliams, A. L., Dean, L. G., Kendal, R. L., & Tennie, C. (2020). Skills and motivations underlying children's cumulative cultural learning: case not closed. *Palgrave Communications*, *6*(1), 1–9. https://doi.org/10.1057/s41599-020-0483-7

Renner, E., Kean, D., Atkinson, M., & Caldwell, C.A. (submitted). Capuchin monkeys' (*Sapajus apella*) and human children's (*Homo sapiens*) use of information from social, virtual, and individual sources in a stimulus choice task.

Sasaki, T., & Biro, D. (2017). Cumulative culture can emerge from collective intelligence in animal groups. *Nature Communications*, 1–6. https://doi.org/10.1038/ncomms15049

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: on the evolution of cumulative culture. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*, 2405–2415. https://doi.org/10.1098/rstb.2009.0052

Tennie, C., Walter, V., Gampe, A., Carpenter, M., & Tomasello, M. (2014). Limitations to the cultural ratchet effect in young children. *Journal of Experimental Child Psychology*, *126*, 152–160. https://doi.org/10.1016/j.jecp.2014.04.006

Terkel, J. (1996). Cultural transmission of feeding behavior in the black rat (Rattus rattus). In C. M. Heyes & B. G. Galef, Jr. (Eds.), *Social learning in animals: The roots of culture*, (pp. 17–47). Academic Press. https://doi.org/10.1016/b978-012273965-1/50003-0

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

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain*

Sciences, *16*, 495. https://doi.org/10.1017/S0140525X0003123X

Wasielewski, H. (2014). Imitation Is Necessary for Cumulative Cultural Evolution in an Unfamiliar, Opaque Task. *Human Nature*, *25*(1), 161–179. https://doi.org/10.1007/s12110-014-9192-5

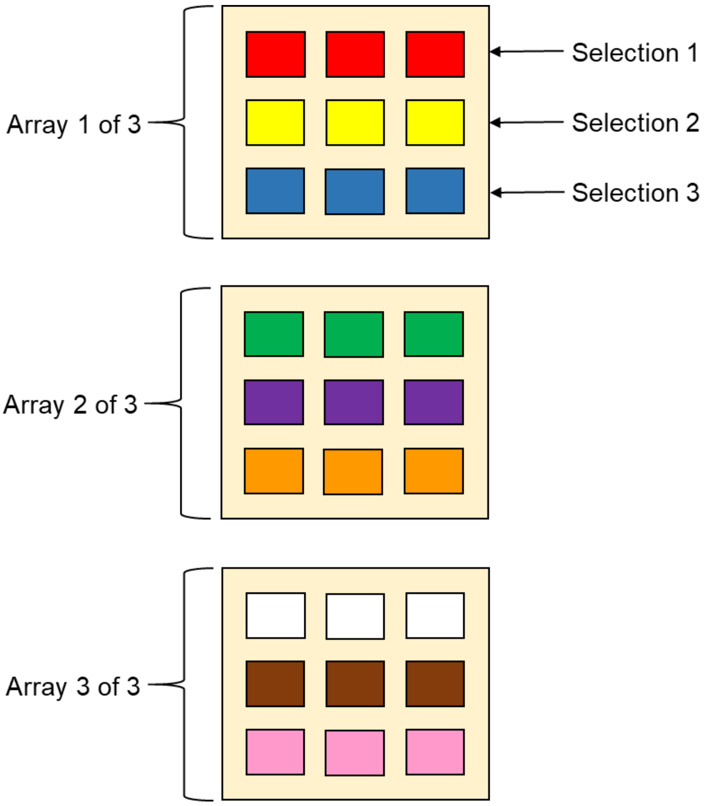
Whiten, A., McGuigan, N., Marshall-Pescini, S., & Hopper, L. M. (2009). Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 2417–2428. https://doi.org/10.1098/rstb.2009.0069

Wynn, T., & Coolidge, F. L. (2011). The Implications of the Working Memory Model for the Evolution of Modern Cognition. *International Journal of Evolutionary Biology*, *2011*, 1–12. https://doi.org/10.4061/2011/741357

Zwirner, E., & Thornton, A. (2015). Cognitive requirements of cumulative culture: Teaching is useful but not essential. *Scientific Reports*, *5*, 1–8. https://doi.org/10.1038/srep16781

Figure 1 (A)

Diagram of Testing Procedures: Arrangement of Chests in Arrays 1, 2 & 3 (Panel A), and Summary of the Testing Procedure for Each Array, Using Array 1 as an Example (Panel B)



Note. Selections 1, 2 & 3 (labelled in panel A) occur for each array.

Figure 1 (B)

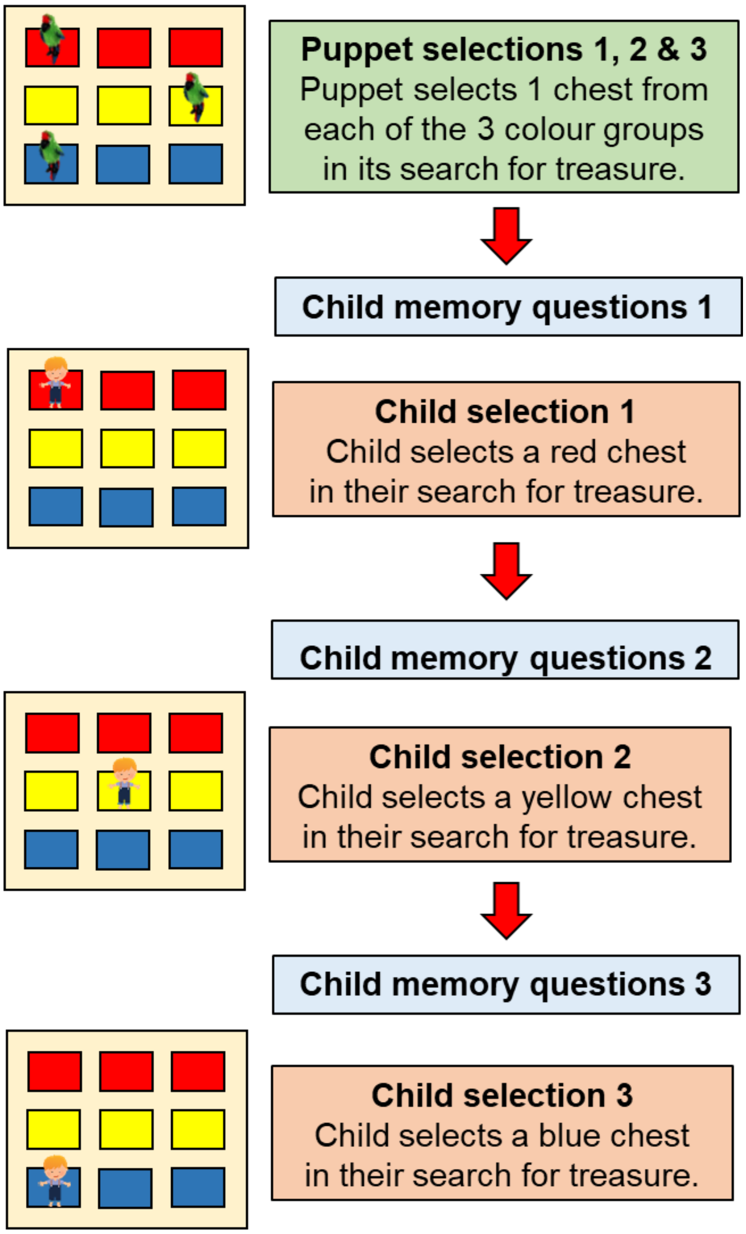
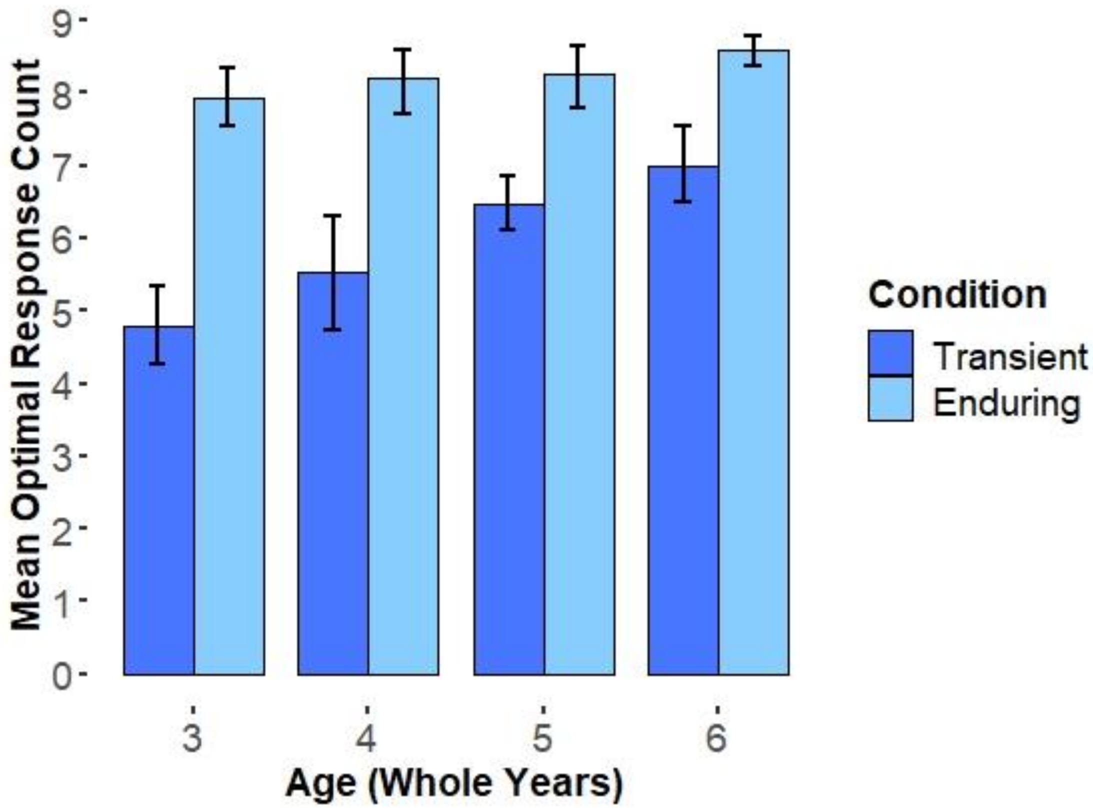

Puppet selections 1, 2 & 3
Puppet selects 1 chest from each of the 3 colour groups in its search for treasure.
Child memory questions 1
Child selection 1
Child selects a red chest in their search for treasure.
Child memory questions 2
Child selection 2
Child selects a yellow chest in their search for treasure.
Child memory questions 3
Child selection 3
Child selects a blue chest in their search for treasure.

Figure 2

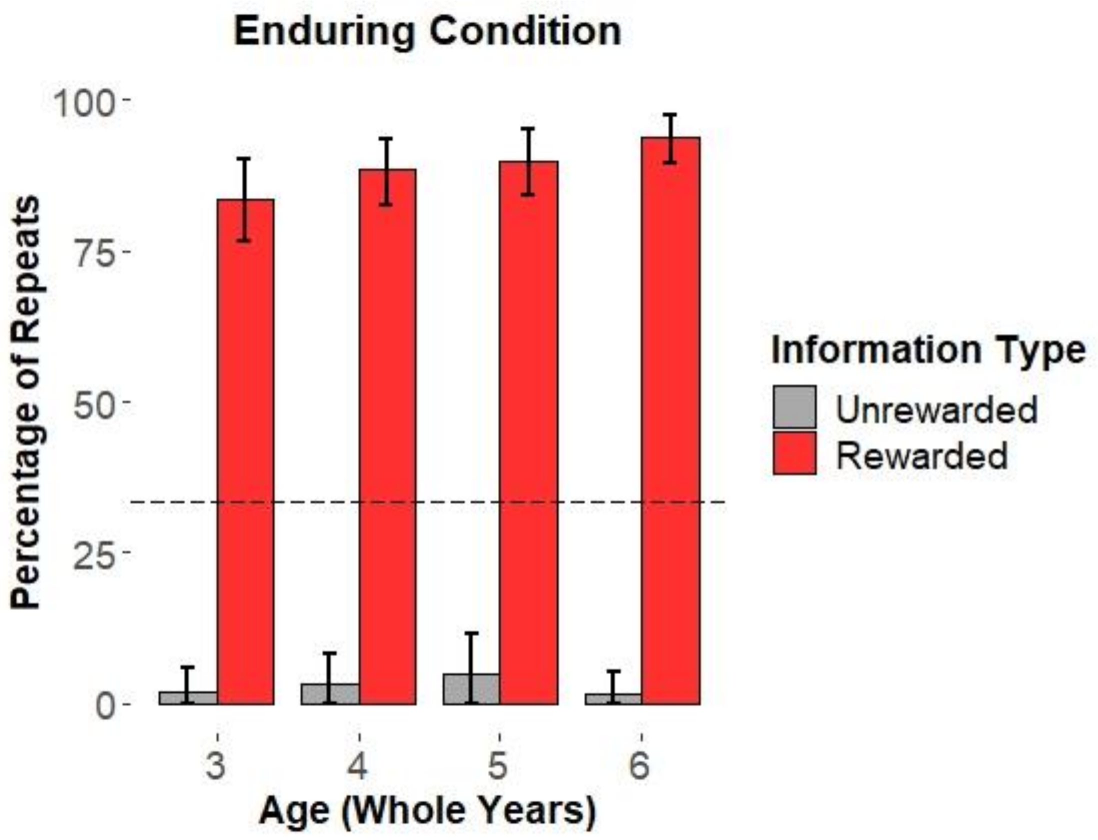
Mean Optimal Response Count/9 by Condition and Age (Whole Years)



Note. Error bars are 95% confidence intervals.

Figure 3 (A)

Mean Percentage of Responses in Which the Location Selected by the Puppet was Repeated for the Enduring Cues (Panel A) and Transient Information (Panel B) Conditions by Information Type and Age (Whole Years)



Note. Error bars are 95% confidence intervals. The dashed line shows chance performance: the proportion of repeats expected (for both unrewarded and rewarded problems) if children were selecting a chest at random on each turn and not using the social information.

Figure 3 (B)

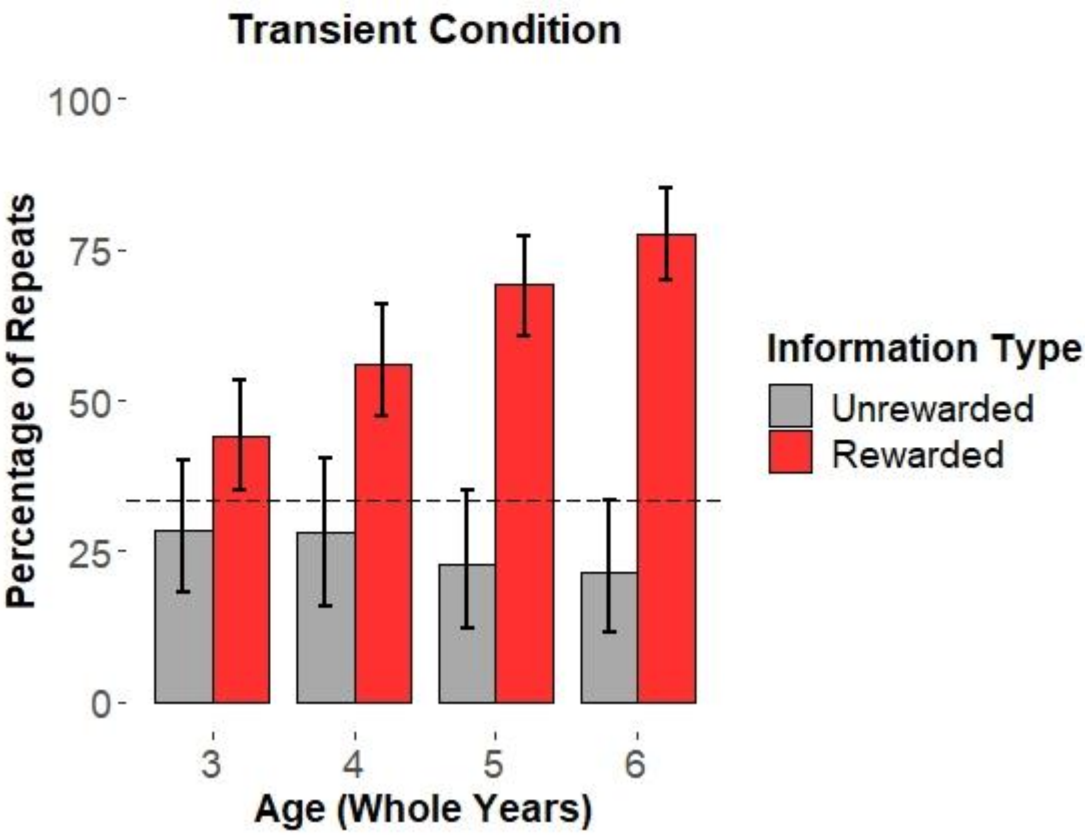
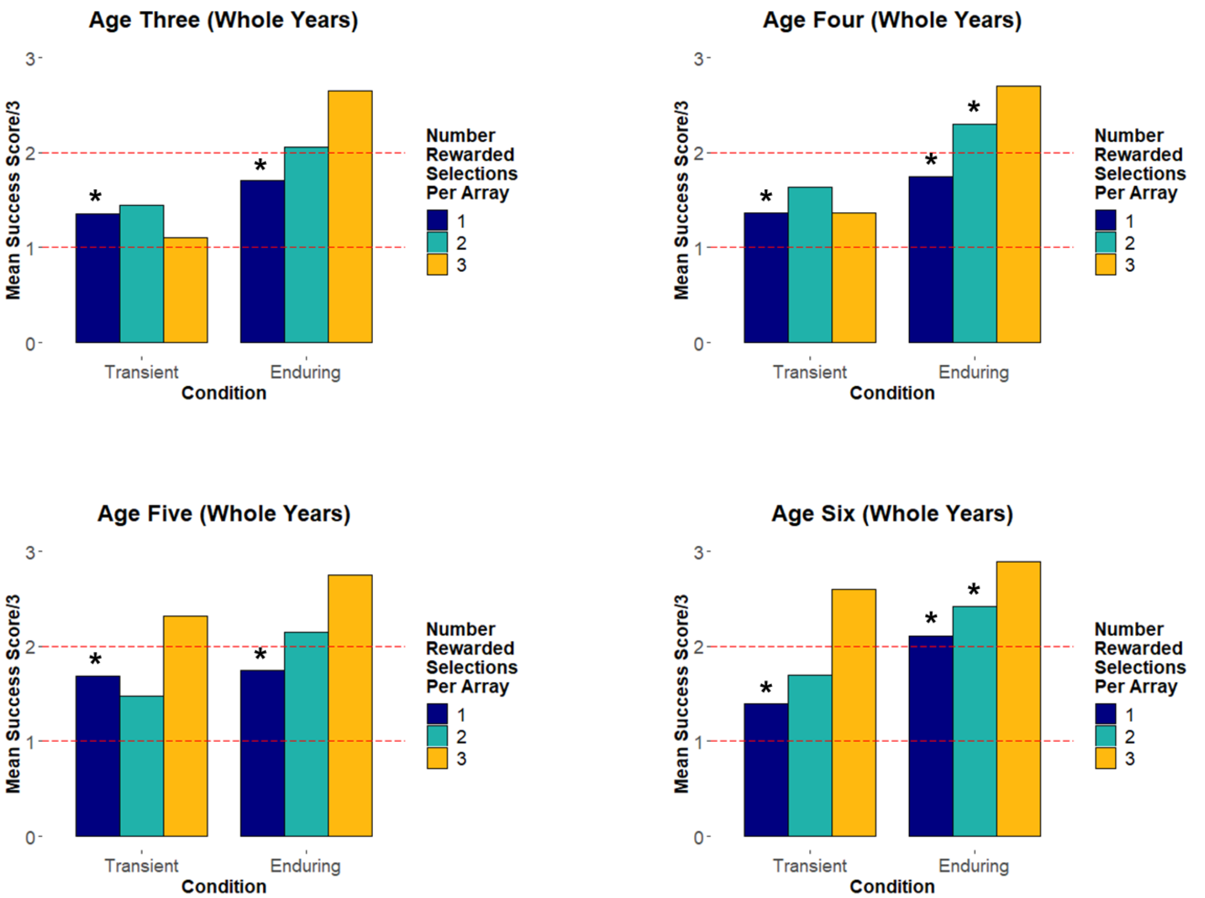

Transient Condition
100
75
50
25
0
Percentage of Repeats
3
4
5
6
Age (Whole Years)
Information Type
Unrewarded
Rewarded

Figure 4

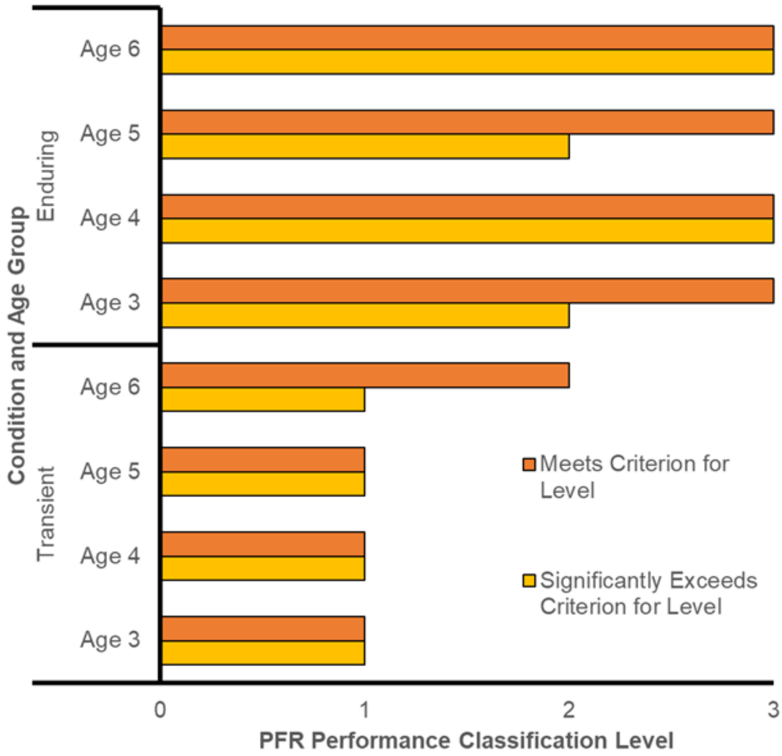
Mean Success Score/3 by Number of Rewarded Selections in an Array and Condition for Ages Three, Four, Five and Six (Whole Years)



Note. The dashed line at 1 depicts chance performance: the score expected if children were selecting a chest at random on each turn and not using the social information. The dashed line at 2 allows visualisation of whether children scored above 2 for the two-rewarded arrays. An asterisk indicates that the mean score is significantly above 1 (chance) or 2 for one- and two-rewarded arrays respectively. An asterisk above both bars denoting one- and two-rewarded arrays, within an age group and condition, illustrates level three ratcheting.

Figure 5

PFR Performance Classification Level by Condition and Age (Whole Years)



Note. Situations in which a mean success score was *significantly* higher (light grey bars) and *higher* but not significantly higher (dark grey bars) than the benchmark outlined in a particular level criterion are described using the terms “Significantly Exceeds Criterion for Level” and “Meets Criterion for Level” respectively.

Table 1

Mean Proportion of Puppet Selections Repeated and Standard Deviation for all Ages Combined and Ages Three-Six (Whole Years) in the Transient Information and Enduring Cues Conditions, Rewarded and Unrewarded Information Types

Age (Whole Years)	Information Type	Condition	Mean Proportion Repeats	SD
All	Unrewarded	Transient	0.25	0.44
All	Unrewarded	Enduring	0.03	0.17
All	Rewarded	Transient	0.62	0.49
All	Rewarded	Enduring	0.89	0.31
3	Unrewarded	Transient	0.28	0.45
4	Unrewarded	Transient	0.28	0.45
5	Unrewarded	Transient	0.23	0.42
6	Unrewarded	Transient	0.22	0.42
3	Unrewarded	Enduring	0.02	0.14
4	Unrewarded	Enduring	0.03	0.18
5	Unrewarded	Enduring	0.05	0.22
6	Unrewarded	Enduring	0.02	0.13
3	Rewarded	Transient	0.44	0.50
4	Rewarded	Transient	0.56	0.50
5	Rewarded	Transient	0.69	0.46
6	Rewarded	Transient	0.78	0.42
3	Rewarded	Enduring	0.83	0.37
4	Rewarded	Enduring	0.88	0.32
5	Rewarded	Enduring	0.90	0.30
6	Rewarded	Enduring	0.94	0.24

Table 2

Mean and Standard Deviation for Success Score/9 and Success Score/3 for Ages Three-Six (Whole Years) in the Transient Information and Enduring Cues Conditions

Age (Whole Years)	Condition	Success Score/9		Success Score/3							
				1 Rewarded		2 Rewarded		3 Rewarded		2 and 3 Rewarded	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
3	**Enduring**	6.41	1.33	1.71	0.92	2.06	0.90	2.65	0.61	2.35	0.52
4	**Enduring**	6.75	1.29	1.75	0.72	2.30	0.66	2.70	0.73	2.50	0.46
5	**Enduring**	6.65	1.35	1.75	0.72	2.15	0.93	2.75	0.55	2.45	0.54
6	**Enduring**	7.42	0.90	2.12	0.74	2.42	0.61	2.89	0.32	2.66	0.34
3	**Transient**	3.90	1.37	1.35	0.81	1.45	0.89	1.10	0.97	1.28	0.55
4	**Transient**	4.37	1.83	1.37	0.76	1.63	0.90	1.37	1.07	1.50	0.75
5	**Transient**	5.47	1.31	1.68	0.82	1.47	1.02	2.32	0.82	1.89	0.59
6	**Transient**	5.70	1.26	1.40	0.60	1.70	0.86	2.60	0.88	2.15	0.49