

RESEARCH ARTICLE



Children's use of social information from multiple models: Cognitive capacities underlying population size effects on cumulative culture

CHARLOTTE E. H. WILKS*, MARK ATKINSON and
CHRISTINE A. CALDWELL

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

Received: December 25, 2020 • Revised manuscript received: November 30, 2021 • Accepted: December 29, 2021

ABSTRACT

Population size has been proposed to promote cumulative culture in humans. Experimental evidence from adult humans suggests that one explanatory mechanism might involve combining beneficial information from multiple models. However, it is possible that such combinatory social learning requires cognitive capacities restricted to adult humans. In our task, children aged 5–10 were exposed to two models who consecutively searched a 3×3 array for rewards. Models revealed different correct and incorrect reward locations. This information could be used by the child to maximise their own score on the same task. We were interested in children's ability to select rewarded locations, and avoid unrewarded ones, revealed by both models. We also manipulated the spatial and temporal displacement of the information available. Results showed that the youngest children were unable to fully benefit from the additional information provided by the two models under spatial and/or temporal displacement. Such displacement likely applies in most real-world cases of cumulative culture therefore our result may offer insight into the constraints on cumulative culture in nonhumans.

KEYWORDS

cumulative culture, population size, cultural evolution, social learning, multiple models, ratchet effect, cognitive development

INTRODUCTION

The human propensity to continuously improve our cultural products over generations of learners has been proposed to be the result of increases in population size (Collard, Ruttle, Buchanan, & O'Brien, 2013; Henrich, 2004; Kline & Boyd, 2010; Kobayashi & Aoki, 2012; Powell, Shennan, & Thomas, 2009; Shennan, 2001). Yet examination of the cognitive capacities which may allow for this population-size effect has been somewhat neglected. It has been suggested that individual learners may utilise the additional information from larger populations through selecting the best of a larger number of individuals from whom to learn (Henrich, 2004), and/or by combining information from multiple different individuals (Kempe & Mesoudi, 2014; Muthukrishna, Shulman, Vasilescu, & Henrich, 2014). However, although such mechanisms are highly plausible, we do not yet know which cognitive capacities underly them and argue that these are likely to be unique to adult humans. If this is the case, then developments in human cognition may form an important part of the explanation as to why increased population size appears linked to vast cultural expansion in humans and not in nonhumans (Dean, Vale, Laland, Flynn, & Kendal, 2014).

Cumulative culture and demography

As humans, our ability to improve and build upon our prior achievements is unparalleled and ubiquitous – evidenced in our languages, complex technologies, societal structures and ability to exploit most environments on the planet. These are examples of what has been

*Corresponding author.
E-mail: charlotte.wilks@warwick.ac.uk

termed cumulative culture, a process whereby a directional pattern of change results in “improvements” (Tennie, Call, & Tomasello, 2009) or increasingly “preferred” traits (Caldwell, 2018) as behaviours or cultural products are transmitted over generations of social learners (Mesoudi & Thornton, 2018). This notion of constant improvement has led to cumulative culture also being referred to as the *ratchet effect* (Tennie et al., 2009; Tomasello, 1990). We use these terms interchangeably. Recent evidence suggests that cumulative culture may exist in nonhumans (Claidière et al., 2014; Jesmer et al., 2018; Sasaki & Biro, 2017; Schofield et al., 2017). However, it is undeniable that this has not constituted anything on the same scale as the phenomena observed across human societies. There are many factors proposed to have driven this cultural expansion in humans (Dean et al., 2014); the factor we focus on here is demography. The results of theoretical models (Henrich, 2004; Kobayashi & Aoki, 2012; Powell et al., 2009; Shennan, 2001), supported by some ethnological studies (Collard et al., 2013; Kline & Boyd, 2010), suggest that changes in population size could have accounted for periods of rapid cultural expansion (e.g. during the upper Palaeolithic 45,000 years ago, Powell et al., 2009; Shennan, 2001) or loss (as in “The Tasmanian Case”, Henrich, 2004) in human groups. Although cited widely, these conclusions have also been fiercely debated (Collard, Vaesen, Cosgrove, & Roebroeks, 2016; Henrich et al., 2016; Vaesen, Collard, Cosgrove, & Roebroeks, 2016) and it is not yet known how changes in population size result in cultural expansion.

Mechanisms underlying the relationship between population size and cumulative culture

Two main mechanisms have been proposed to enable the exploitation of information from larger populations, leading to greater cultural complexity. Firstly, success-biased copying, whereby learners selectively copy the individual with the best available (randomly derived) variant of a cultural trait (see Wood, Kendal, & Flynn, 2013, for a review of such biased social learning strategies). This mechanism has been exemplified in theoretical models (e.g. Henrich, 2004; Powell et al., 2009; Shennan, 2001) which have investigated conditions under which larger numbers of interacting individuals (“effective” populations) can support behaviours or products of increased cultural complexity/functionality. For example, in Henrich’s (2004) model (extended by Kobayashi & Aoki, 2012) learners engaged in such success-biased copying but the copying was inaccurate which would, under certain conditions, ultimately lead to a loss of skill within the population. However, larger populations had an advantage: they were able to negate this loss because they contained more individuals with traits which were of higher than average fitness, hence a learner was more likely to select one of these more successful individuals from whom to learn. This was found to counteract the negative effects of copying (low transmission fidelity), leading to cumulative culture in populations above a certain size threshold. Derex, Beugin, Godelle, and Raymond (2013) performed experimental work

which supports the validity of success biased copying to exploit information from larger populations, as proposed by these models. The authors introduced a simple and a complex artefact building task into groups containing two, four, eight or 16 members (players in a computer game) and on each of multiple rounds players could select to learn from just one player from the previous generation. Overall performance on both tasks increased with group size. Moreover, only the larger two groups eventually scored more highly than an initial demonstration in the simple task and avoided deterioration from this benchmark in the complex task.

Secondly, based on further modelling work, some authors have proposed that combining information from multiple different models can lead to improved trait variants (Lewis & Laland, 2012), as required for cumulative culture. Moreover, innovation through combination has been proposed to be an important mechanism underlying the evolution of technology (see Winters, 2020, for examination of the importance of this mechanism versus minimisation of information loss through the use of social learning mechanisms). Evidence consistent with this has been identified in the form of phylogenetic analyses on the constituent elements of technologies; existing forms of devices, such as radios (O’Brien & Lyman, 2000) or bikes (Lake & Venti, 2009), have been combined to produce the latest versions (Muthukrishna & Henrich, 2016). Furthermore, there is now experimental evidence (Kempe & Mesoudi, 2014; Muthukrishna et al., 2014) that adult learners with access to multiple models integrate information across models to generate better solutions than those with access to one model.

Muthukrishna et al. (2014) tasked participants in transmission chains to recreate a complex image using image editing software. The transmission chains were ten-generations long and were structured such that participants were exposed to information from either one or five models from the previous generation. Each participant provided a screenshot and written instructions to help those in the subsequent generation. There was evidence of cumulative culture in the five-model group only, and furthermore, participants appeared to preferentially utilise information from the top performing model, in addition to the next three top performers. This provided evidence that participants combined information from multiple models to generate novel combinations possessed by none of their cultural parents in addition to employing success biased copying.

Kempe and Mesoudi (2014) ran transmission chains with either one or three models in each of four generations. The task was to complete a 100-piece jigsaw puzzle and participants had full access to the attempts of the participant(s) from the previous generation. As in Muthukrishna et al. (2014), there was evidence of ratcheting in the group with multiple models only (comprising five separate chains), measured as an increase in the mean number of puzzle pieces correctly connected to other pieces as generation increased. It appears that this ratcheting was possible because the presence of multiple models provided an increased amount of information which could be integrated. This was apparent from the number of unique puzzle pieces

correctly connected by all three models within a single generation for any given chain, i.e. the total package of information about the puzzle which was passed to the next generation. Increasingly therefore, theorists have argued that the production of improved variants (that is, trait variants which represent an improvement or upgrade on those previously found in a population) through novel invention may play a relatively insignificant role in advancing cumulative culture (Lewis & Laland, 2012), and that the popular supposition of “genius” inventors should perhaps be reconsidered (Muthukrishna & Henrich, 2016). Nevertheless, combining information from multiple models inevitably involves an element of individual learning. When we discuss an individual combining (or integrating) information from multiple models (or pre-existing technologies) we are referring to a situation in which information from multiple models is utilised, but not necessarily one which is absent of input from the individual.

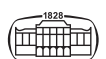
Cognitive demands associated with utilising information from multiple models

Understanding the above mechanisms can shed light on the relationship between demography and cumulative culture. Although the mechanisms themselves may be used by both humans and nonhumans, as the amount of potentially useful information increases possession of certain cognitive abilities by adult humans (discussed further below) may be of increasing benefit in order to deal with demands associated with using the information. Our previous work (Wilks et al., 2021) is relevant here. We examined whether the propensity to show cumulative culture is dependent on both the context in which the information is presented and cognitive ability, and proposed that humans are likely to show cumulative culture in a greater range of contexts than nonhumans due to enhanced cognitive abilities. We investigated whether children of different ages (3–6 years-old), and thus working memory capacities, were able to improve on information provided by one model under two different task contexts in which the presentation of the information (reward locations in a treasure-hunting game) was manipulated. One context was more cognitively taxing on memory – sought-after reward locations revealed by the model were masked and needed to be held in memory prior to the child searching the same space for rewards. In the other context the information from the model’s search remained visible. We found support for our hypothesis; in the less challenging condition children of all ages were able to find more rewards than the model and show the potential for cumulative culture, whereas in the condition which taxed memory only the 6-year-olds showed this potential. This study focussed on an ecologically valid task context in which information was provided by *one* model. However, the study also speaks to the use of information from multiple models because we showed that the way in which information is presented in a social learning context can affect use of this information, the cognitive capacities required to use it, and thus the potential for cumulative culture. Using information from multiple models

may present similar, and likely greater, inherent challenges due to the way in which the information is presented and the greater amount of information available for use. We discuss these potential challenges below.

If cognitive abilities unique to humans, or used differently in humans, are implicated in overcoming the challenges associated with making use of information from multiple models then this may contribute to the explanation as to why increases in population size appear linked to the expansion of cumulative culture in humans. Yet this theory has received little academic attention. In the present study, we therefore investigate children’s ability to utilise information from multiple models under different cognitive constraints. We included experimental constraints which were designed to reflect inherent difficulties in utilising information presented by multiple models. Before discussing these in detail, we briefly present some literature which supports the idea that using information from multiple models is inherently cognitively demanding, such that benefits cannot be guaranteed under all circumstances, and may even be *reduced* by the increased cognitive burden.

Caldwell and Millen (2010) and Fay, De Kleine, Walker, and Caldwell (2019) investigated the relationship between population size and cumulative culture using transmission chains in which the participants’ goal was to build paper aeroplanes which could fly as far as possible. Both studies included conditions in which participants had access to information from different numbers of models (one, two or three in Caldwell & Millen, 2010, and one, two or four in Fay et al., 2019); the study by Fay et al. (2019) also included an individual learning condition (repeated attempts replacing generational turnover). Rather surprisingly, Caldwell and Millen (2010) found cumulative improvements in plane flight distance as generation number increased for the one and two model conditions, but not the three model. Similarly, in Fay et al. (2019) flight distance increased with generation number for the individual learning and one model conditions, but not the two and four model conditions. In Fay et al. (2019) the previous generations’ planes (e.g. four in the four-model condition) were only available to view (individually) for a short period prior to a participant attempting to build their own plane. In Caldwell and Millen (2010), although some of the previously built planes were on display during a participant’s building time (e.g. three in the three-model condition), the time available to view each individual plane in the three-model condition was extremely short, with each removed after only a brief exposure time. Participants therefore did not have continuous access to all of this potentially valuable information for the full building period. In these studies, it therefore seems probable that processing the different models’ designs, within a restricted time period, and keeping them in working memory in order to select the best design (or make use of the beneficial elements from previously observed designs) became more challenging as the number of models increased. That is, once the number of models increased beyond a certain point, cognitive demands associated with integrating the information constrained its usefulness. These studies therefore



support the theory that using social information from multiple models may pose significant cognitive challenges due to the increased information and the way in which it is presented; we now discuss this further.

Useful information from multiple models may be separated in time and space from a potential social learner's own attempt, and we suggest that exploiting this information is likely to involve increased cognitive load (Caldwell & Millen, 2010; Fay et al., 2019). For example, potentially beneficial information from multiple models may be observed at different time points and remembering and storing this information for later use would be cognitively demanding on memory. Furthermore, information provided by multiple models may contain a range of elements, some of which the information user wishes to utilise and others they wish to discount. Holding in mind these different elements, whilst deciding how they are best utilised, may involve working memory (defined as both the storage and manipulation of information, Best & Miller, 2010; Cowan, 2008; Diamond, 2013; Garon, Bryson, & Smith, 2008). Consider a "real-life" social learning scenario – integration across time might be required for learning a foraging skill and could involve combining information from observation of an individual extracting the contents of a shelled plant, with information obtained through a separate observation of the technique required to break into this item.

Alternatively, it may be challenging for a learner to recognise the value of information from other individuals in relation to oneself (Blakey et al., 2020) and to bring this information together for use. For example, other individuals could be interacting with an object equivalent to one currently available in the learner's own immediate vicinity. Using this information will likely involve spatially translating what they have observed to their own bodily frame of reference. This probably requires skills in recognition and mapping of correspondences (DeLoache, 1989, 1991, 2000), and mental translation and rotation (Frick, Hansen, & Newcombe, 2013; Levine, Huttenlocher, Taylor, & Langrock, 1999; Shepard & Metzler, 1971). To use the same example as above, integration over space (but not time) would be involved if the learner (with access to their own fruit or nut, allowing simultaneous activity) could observe two individuals concurrently, one of whom was demonstrating the opening technique, and the other the extraction.

Both temporal and spatial displacement of information therefore generate increased cognitive load for a potential information user, with demands on memory, spatial translation and evaluation of alternative information. Moreover, we would expect this cognitive load to be further increased if the information to be remembered, mapped or translated comes from more than one source. It follows that a learner with a more limited capacity for domain-general cognitive processing would be less able to exploit information presented by multiple models in comparison to adult humans; we expect this to include young children and nonhumans. However, we expect children to become better at utilising information from multiple models with increasing age as

capacities such as working memory and metacognition develop. We investigated this experimentally, across a broad developmental range, in conditions in which the information from multiple models was presented differently in time and/or space.

Children's ability to integrate information from multiple models

Research exploring children's ability to integrate information from multiple models is in its infancy, and not directly comparable to the adult experimental work previously discussed. Nevertheless, we outline three studies which are a relevant starting point, before introducing our study in greater depth. Subiaul, Krajkowski, Price, and Etz (2015) studied 3–5-year-olds' ability to open a two-compartment box when provided with adequate demonstrations from one model (opening both compartments), two models (each opening one compartment) or in an individual learning condition with no demonstrations. They found that children were more likely to successfully open both compartments in the model conditions compared to the individual learning condition, but that there was no difference in performance between the one and two model conditions. This is not surprising, because the actual information provided in the one and two model conditions was the same – the demonstration from the two models as a whole was identical to that provided by one model. This study therefore showed that the presence of two models did not present a problem in terms of using the available information. However, unlike in the adult experimental work, in this study using the information from two models does not require any of the realistic constraints which might operate when integrating information from multiple models in the real world, such as separation of the information in time (Fay et al., 2019) and/or space (Kempe & Mesoudi, 2014; Muthukrishna et al., 2014). Additionally, this study did not require children to be selective regarding the information they used from each model – each provided exactly half of that in the full, one model, demonstration. Thus, simply summing the information across the two models would result in perfect performance. In real-world cases of cumulative culture, learners are more likely to be exposed to multiple, imperfect demonstrations which together may contain all the information needed, but which also contain potentially distracting information about other behaviours not linked to success. The selective extraction of this relevant information, and selective inhibition of any redundant or ineffective elements, probably brings with it significant cognitive challenges (as discussed above) that were not captured within the design of Subiaul et al. (2015).

However, a more recent study by Subiaul and Stanton (2020) presents a closer representation of such real-world challenges. The authors demonstrated that children and adults were able to combine distinct information (two sections of a tower, each made of two plastic pieces) presented by two models to produce a new, optimal solution (the tallest tower possible from four plastic pieces). The tower



formations produced by the models were dismantled following presentation, introducing a temporal separation between demonstration and potential use of the information. This required children, in what the authors termed “summative learning groups”, to remember these distinct tower formations in order to spontaneously reproduce and combine them. Although this introduced an additional, and ecologically valid, cognitive challenge it did not involve selective extraction of the most useful information because simply combining the two demonstrations resulted in the most optimal tower. An experiment by Buchsbaum, Gopnik, Griffiths, and Shafto (2011) demonstrated that 3–5-year-olds do have the ability to use information selectively, although this study did not look specifically at learning from multiple models. Children were shown five different action sequences (which caused a toy to play music) and were able to integrate this information to eliminate actions which were causally irrelevant. However, this study did not require children to integrate *successful* elements across different demonstrations. Nevertheless, it does demonstrate that children have the ability to integrate information selectively across multiple demonstrations.

The developmental literature outlined above demonstrates that children can use information from multiple models in principle, and that they can integrate simple information selectively. However, there is currently no evidence that they can selectively integrate successful elements from multiple models to the extent that an improved solution (improved trait variant) is introduced into an experimental population. Consequently, further work is required before we can draw conclusions as to children’s ability to use information from multiple models to generate examples of cumulative culture.

The present experiment

Building on the results of Subiaul et al. (2015) and Subiaul and Stanton (2020), in addition to the aforementioned adult studies (Kempe & Mesoudi, 2014; Muthukrishna et al., 2014), we explored children’s ability to extract relevant information, and eliminate that which was ineffective, to perform more highly than the highest performing single model in a two-model population (thus creating an improved trait variant). Furthermore, we investigated how this ability changed during development (across ages 5–10) and as the information provided by the two models differed in its temporal and/or spatial displacement. This would enable us to examine whether human-unique cognitive processes (such as those outlined below), or the way in which humans make use of these, may be necessary for this kind of information use. We did not test children below age 5 because our previous work using a similar task, which included a condition taxing working memory, had shown that until age 6 children found using information from even just one model extremely difficult (Wilks et al., 2021). We thus predicted that children below age 5 would find certain conditions in our current task even more challenging, but that testing children from age 5 to 10 should provide

sufficient scope to document age-related effects on task performance.

The information from the two models was presented to children in the form of a searching task (adapted from our previously successful paradigm; Wilks et al., 2021) presented on a touchscreen computer, in which both models individually searched a space for hidden treasure. In the current study, the rewards (“treasure” – gold coins) were always hidden in three of nine chests presented as 3x3 arrays on each of 24 trials. Each model made three selections from each nine-chest array, so the maximum number of reward locations which could theoretically be revealed, both by each model and collectively across the two models on each array, was three. We placed constraints on the selections made by both models; these conformed to six different demonstration types (Table 1), each of which was presented once within each of four conditions (described below). These demonstration types varied in relation to the highest score achieved by a single model in the pair (equivalent to the number of rewards found by this model – 1 or 2), the mean score across the two models, and the total number of rewards found (out of 3) by both models combined (see Table 1).

We expected the least challenging of our four conditions to be our “Increased Information” condition in which the information provided by the models was not displaced either temporally or spatially. That is, both models sequentially searched the array which the participant themselves would later search. In addition, both models’ selections remained visible in the search space whilst the participant made their own selections. This condition acted as our control because the information provided was effectively the same (at least in terms of the cognitive challenges involved) as if it had been provided by a single model. However, the condition also controlled for any distraction or facilitation caused by the presence of two models. In contrast, our “Temporal” displacement condition presented a greater cognitive challenge. Although the models once again made their selections from the same array, the locations selected by each model were concealed again, immediately after the model’s third selection. Therefore, in order to fully integrate the information, the locations and reward value (rewarded/unrewarded) needed to be held in memory. We also included a “Spatial” displacement condition in which the first model made their selections from an array on the top left of the screen, the second model from an identical array on the top right, and the child from another such identical array on the bottom middle of the screen. The rewards were hidden in the same positions across the three arrays, but we were interested in whether children could integrate useful information from the top left and right arrays. As in the Increased Information condition, the reward locations selected by the models remained visible throughout the child’s selections. Our “Temporal-Spatial” displacement condition was expected to be the most cognitively challenging because it combined the temporal and spatial elements from the Temporal and Spatial conditions respectively. That is, the models made their selections on two arrays, as in the Spatial condition, but the revealed

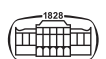




Table 1. Summary of the Six Demonstration Types Denoting the Models' Selections (Each Presented Once Within the Increased Information, Temporal, Spatial and Temporal-Spatial Conditions)

Highest single model score	Demo type	Number rewarded (R) & unrewarded (U) selections –model 1	Number rewarded (R) & unrewarded (U) selections –model 2	Mean score across the two models	Expected chance score	Total number of rewarded locations revealed	Total number of unrewarded locations revealed	Number of rewarded locations selected by both models (repeated)	Number of unrewarded locations selected by both models (repeated)	Number of rewarded locations selected by one model only	Number of unrewarded locations selected by one model only	Number of locations with unknown contents following models' selections
1	3	R UU	R UU	1.00	1.00	2	3	0	1	2	2	4 (1R, 3U)
1	5	R UU	UUU	0.50	1.00	1	3	0	2	1	1	5 (2R, 3U)
1	6	R UU	R UU	1.00	1.00	1	3	1	1	0	2	5 (2R, 3U)
1	Mean 3,5,6			0.83	1.00							
2	1	RR U	R UU	1.50	1.00	3	2	0	1	3	1	4 (4U)
2	2	RR U	RR U	2.00	1.00	3	2	1	0	2	2	4 (4U)
2	4	RR U	R UU	1.50	1.00	2	2	1	1	1	1	5 (1R, 4U)
2	Mean 1,2,4			1.67	1.00							

locations were concealed again immediately after each model made their third selection, as in the Temporal condition. To utilise the information fully therefore required both an ability to integrate information from two spatially distinct locations and the capacity to hold this in one's working memory.

Our expectation was that children would utilise the information from the two models across a wider range of conditions as their ability to overcome additional cognitive demands increased with age. Moreover, we expected that children's score, and ability to repeat the revealed rewarded information (and not repeat the unrewarded), would increase with age in the Spatial, Temporal and Temporal-Spatial conditions, which presented constraints on working memory and spatial integration; we did not expect these to increase with age when such constraints were absent in the Increased Information condition. Our experimental design does not allow us to accurately pinpoint the exact, developing cognitive capacities which may be involved in utilising information from multiple models, and providing a definite answer to this question is beyond the scope of this paper. However, investigating whether there are performance improvements in contexts in which the information is separated over time/space has the potential to be an informative first step.

Predictions

For example, if (as predicted) performance improves with age in the Temporal and Temporal-Spatial conditions in particular (whilst remaining static in the control condition) this could implicate that working memory, and/or enhancing working memory through metacognitive processes, aids use of information from multiple models. Working memory continues to increase in a gradual, linear fashion from pre-school age through to the teenage years (Best & Miller, 2010). However the picture is complicated by the use of tasks which vary in difficulty. Less complex tasks require the holding of information in the mind for a short time without manipulation (for example the forward digit span – Woods et al., 2011), whereas in the more complex tasks maintenance and manipulation of the information is required to achieve success (an example task being the backwards digit span – Best & Miller, 2010; Garon et al., 2008). The constraints involved in our task are more reminiscent of those presented by complex tasks, for which performance continues to improve up until age 5, with some authors showing improvement continuing until age 7 (Garon et al., 2008) or even into adolescence (Best & Miller, 2010). Therefore, we might expect to find continued improvement in our task between ages 7 and 10 in the conditions involving temporal displacement.

In terms of metacognition, we would expect involvement to result in increased task success (across all three displacement conditions) from age 6 as children develop their skills in a range of metacognitive processes. For example, there are significant advancements in children's metacognitive understanding of how perceptual access leads

to knowledge at around age 6. Studies have shown that prior to this age children have difficulty understanding that partial perceptual information leads to incomplete knowledge and children therefore frequently overestimate their own knowledge (Kloo, Rohwer, & Perner, 2017; Rohwer, Kloo, & Perner, 2012; Sodian & Wimmer, 1987) and that of others (Chandler & Helm, 1984; Taylor, 1988). Accurately assessing one's own and others' knowledge may be important in order to take advantage of the most useful information from multiple models. Furthermore, children aged 6 plus, with more experience in formal schooling (Bryce & Whitebread, 2012), are likely better equipped to utilise metacognitive storage processes such as mnemonic devices (Jurowski et al., 2015) or inner speech (Carruthers, 2013; Cowan, 2008) to lessen cognitive load, and in particular memory load. Such strategies have been described as "cognitive offloading" and children's tendency to use (Armitage, Bulley, & Redshaw, 2020) and devise (Bulley, McCarthy, Gilbert, Suddendorf, & Redshaw, 2020) these increases with age. Children may also show increased metacognitive monitoring (e.g. judging task difficulty, self-questioning) and control (e.g. planning, changing strategy) with age (Bryce & Whitebread, 2012).

Measuring the potential for cumulative culture

In order to investigate children's ability to produce an improved trait variant, and hence potential for cumulative culture, we grouped according to chronological age, using one-year bandings, and split the data further depending on condition (Increased Information, Temporal, Spatial or Temporal-Spatial) and whether the highest single model score in a given trial was 1 (henceforth model-score-1 trials) or 2 (model-score-2 trials). We also calculated the mean child score for each age group, condition, and model-score groups (1 or 2). For the model-score-2 trials only, we planned to analyse whether these mean scores were significantly greater than the highest single model score of 2. Outperforming the highest observed single model score was analogous to outperformance of the best available information within a population, an improved variant. Additionally, again for the model-score-2 trials only, we were interested in whether the mean child scores were significantly greater than the mean score across the two models. Outperforming the mean score of the two models could be used as a proxy for outperformance of a randomly selected single model. Linking back again to our interest in cumulative culture, this measure would give an indication of whether children could outperform the previous generation's "typical" score. This is important because it would show whether, *on average*, later generations would accumulate benefits relative to their predecessors. We also measured whether children were scoring significantly above chance level (i.e. the score expected to arise from naive exploration in the absence of any information) for both the model-score-1 and 2 trials. A chance level score was 1 because each child selected three out of nine chests on each array and there was a total of three rewarded chests within each array. This comparison with chance would serve to test



whether the children were indeed using some of the information available to them in the demonstration (even if not optimally). This is because scores not significantly different from chance could not be distinguished from a pattern of random selections, uninfluenced by the demonstration information. In the model-score-1 trials this chance-level of 1 was equivalent to the highest single model score. Therefore, for the model-score-1 trials, scoring significantly higher than the highest single model score was effectively captured by this same comparison. Furthermore, in the model-score-1 trials, the mean score across the two models (0.83) was lower than the chance-level therefore outperformance of the mean did not provide us with meaningful information, over and above the comparison to chance, for the same reason. The benchmarks of outperforming the highest single model score and mean score, as used for the model-score-2 trials, were therefore not meaningful for the model-score-1 trials.

MATERIALS AND METHODS

Participants

169 children were recruited from a primary school in Stirling, Glasgow Science Centre, and a science festival attended by members of the public at the University of Stirling. Five children (all male) were excluded due to: missing date of birth from the consent form meaning age could not be confirmed and age in days could not be calculated (age 8), failure to fully comply with task instructions (age 10), a recognised developmental delay ($n = 2$, ages 6 & 7) or because they were found to have participated previously (age 6). The final sample consisted of 164 children aged 5 to 10 ($M = 7$ years, 10 months; range = 5 years, 0 months–10 years, 11 months; $SD = 1$ year, 8 months; 79 female); there were between 20 and 31 children in each age group.

Experiment

The task was presented to children on either a touchscreen laptop or tablet running Windows 10 and was written and

run in PsychoPy, version 1.84.2 (Peirce et al., 2019). Task responses were automatically written into a csv file and any verbal comments children made were recorded on paper by a research assistant.

The goal of the task was to find as many pieces of treasure (gold coins) as possible following watching two social models attempt the same task. Each of 24 experimental trials (plus four practice trials) contained either one (Increased Information and Temporal displacement conditions, Fig. 1, left) or three (Spatial displacement and Temporal-Spatial displacement conditions, Fig. 1, right) 3x3 array(s) of nine treasure chests, three chests of which were rewarded and six unrewarded. Rewarded chests revealed a gold coin, accompanied by the sound of money, if selected, and unrewarded chests a red cross and a beep. The experiment used a within-subjects design. Participants took part in four experimental conditions, which varied in terms of whether the information provided by the two models was separated from their own attempt, either in time or space (Increased Information, Temporal displacement, Spatial displacement and Temporal-Spatial displacement, see introduction). Note, the two conditions involving separation in space (Spatial and Temporal-Spatial displacement) also differed from the other conditions (Increased Information and Temporal displacement) in that the models made their selections from two different search spaces as opposed to one, introducing an additional element of spatial separation. As stated in our introduction, we placed constraints on the selections made by both models which conformed to six different demonstration types (further details in Table 1).

To determine the order of the conditions, we randomly selected from the 24 different permutations of the four conditions (Increased Information, Temporal, Spatial and Temporal-Spatial) for each participant. The two social models were a cartoon parrot “Pirate Parrot” and octopus “Pirate Octopus”. Between participants, we counterbalanced which model made their selections first. Therefore, the model choosing first remained consistent for a participant throughout the task. The order of the six demonstration types was randomised for each participant within each condition. Within a demonstration type, we also randomised

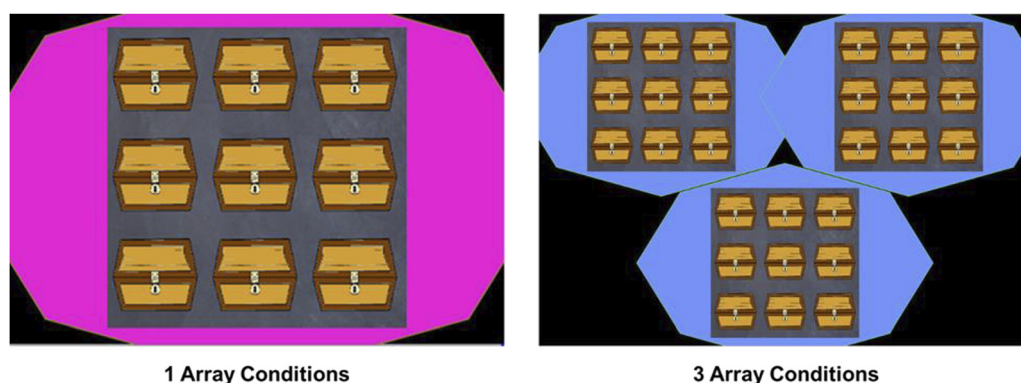


Fig. 1. Example task display containing one (left image) or three (right image) arrays as in the Increased Information/Temporal and Spatial/Temporal-Spatial conditions respectively

whether the parrot or the octopus performed the selections designated for model 1 or model 2 and the actual locations selected within the array.

Demonstration types. Across the six demonstration types (1–6, Table 1), the highest single model score over the two models on any given trial was either 1 or 2 (three of each type); the mean score across the two models varied between 0.5 and 2.0; the total number of rewards revealed across the two models combined was either one, two or three (two of each type); and the number of rewarded selections repeated by each model within an array was either 1 or 0 (three of each type). We did not include a demonstration with zero rewards because we wanted to ensure that children were receiving some information that would be of benefit to copy on each trial. We did not include demonstration types in which any one model found all three rewards because we were interested in children's ability to use information from the two models to find more rewards than the highest performing model. We were also interested in 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).

Procedure

In the school, testing took place in a quiet area adjacent to the classroom. At Glasgow Science Centre and a science festival held at the University of Stirling, testing was conducted in a public space, separated from the main museum/festival space by a desk or room partition. At the science centre and festival 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 assistance relevant to the task. In each location the laptop or tablet running the task was positioned on a table-top, which the child sat in front of at a comfortable distance; the experimenter sat next to the child. A verbal script was used by the experimenter (see SI) and total testing duration was 15–20 min per child.

Introduction to task. Children were asked if they would like to play a game in which the goal was to search for treasure. They were first shown a series of on-screen images (Fig. 2), supported by verbal instruction and explanation from the experimenter. The gold coins, arrays of treasure chests, notion of rewarded/unrewarded chests and the models were all introduced. Children were informed that the models would also be looking for treasure and that they should watch them carefully before taking their turn. The goal of the experiment, and the fact that they should try to find all three pieces of treasure in each array, were reinforced several times throughout the introduction.

Practice trials (x1 per condition). Following the introduction, a practice trial corresponding to the first assigned condition was conducted prior to the six experimental trials for that condition. The six experimental trials covered each of the six demonstration types (Table 1). The remaining

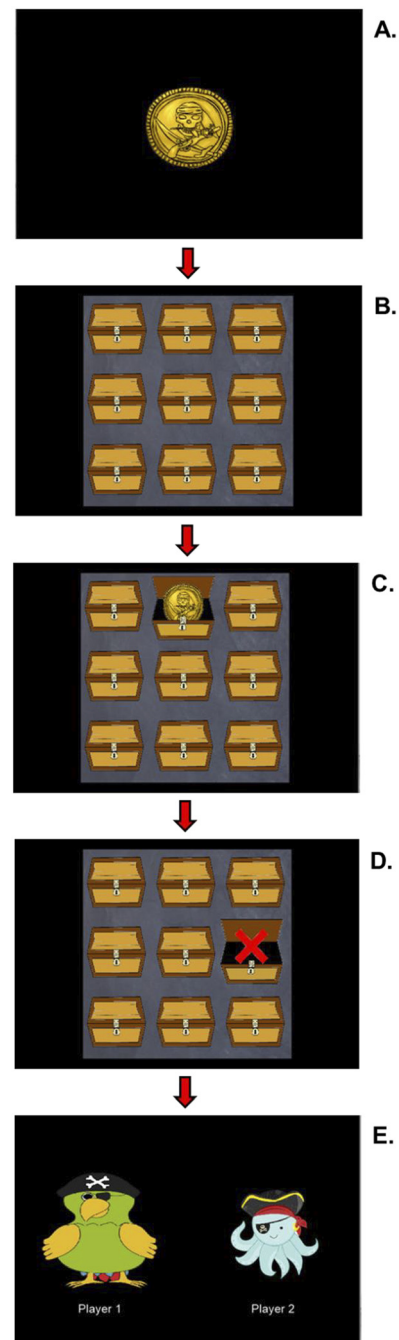


Fig. 2. Images displayed to children during the task introduction

three conditions were run in the same way, with a practice trial followed by the six experimental trials.

A practice trial began with the presentation of a single array set back slightly from the forefront, in the middle of the screen. This array then moved forward for the single array conditions (Increased Information and Temporal) or gave the illusion that it was splitting into three for the three array conditions (Temporal-Spatial and Spatial). In the latter case the top half of the screen contained one array on the left and one on the right, and the bottom half one array in the middle (Fig. 1, right image). However, the rewards were in the same location in each of the three arrays. The first of the

two models then selected three chests (see Fig. 3A for an example for the Spatial condition) and upon each selection a chest opened to reveal either a coin (rewarded) or a red cross (unrewarded) and accompanying sounds. Immediately following the third selection chests either closed completely, hiding their content, or partially, meaning that their content remained visible. In the conditions with three arrays this first model's selection was made on the top, left array (Fig. 3A). The second model then proceeded to select three chests in the same manner as the first. In the single-array conditions this was on the same array as the first model and in the three array conditions the top, right array (Fig. 3B). All selected chests then either closed completely, or partially (Fig. 3C). Next, the child was prompted to search for all three pieces of treasure. It was made clear that they could select a chest even if one, or both, of the models had already selected it, as we wanted to ensure (in the Increased Information and Spatial conditions especially) that children understood that a model selecting a correct chest did not mean that the chest had been emptied. Children made their selections by touching three of the nine chests in turn on either the same array as the two models (one array conditions), or the array in the bottom, middle of the screen (three array conditions), Fig. 3D. As occurred for the models' selections, the chests initially opened to reveal their contents as they were chosen, but after the third selection the three chests remained open and the remaining six, unselected chests, also opened so that the contents of all nine chests were visible (Fig. 3E). The experimenter then stated "Look – that's where the treasure was" and pointed to the rewarded chests in turn in order to reinforce that exactly three of the nine chests were rewarded on every trial. The child was asked: "How many pieces of treasure were there?" This question assessed understanding and if a child answered correctly or incorrectly their response was reinforced or corrected respectively.

Experimental trials (x 6 per condition). The six experimental trials (spanning the six different demonstration types – Table 1) followed the practice trial for that condition and were almost identical except that there was less intervention from the experimenter. The participant was not reminded that they were searching for three pieces of treasure and was not asked how many pieces of treasure were revealed following their selections, although the experimenter still pointed to each of the three pieces of treasure when all chests opened at the end of the trial. The only other difference was that the single or three arrays (Fig. 3) appeared immediately at the start of the trial and did not need to be moved forward (single array condition) or give the illusion of splitting into three (three array condition).

Ending of experiment. Following completion of all four conditions (each consisting of a practice trial followed by six experimental trials) a black screen was presented. A large, gold coin superimposed onto a chest appeared in the middle of the screen and a number above the chest continued to increase incrementally until it represented the total number of rewards found by the participant – a maximum of 72

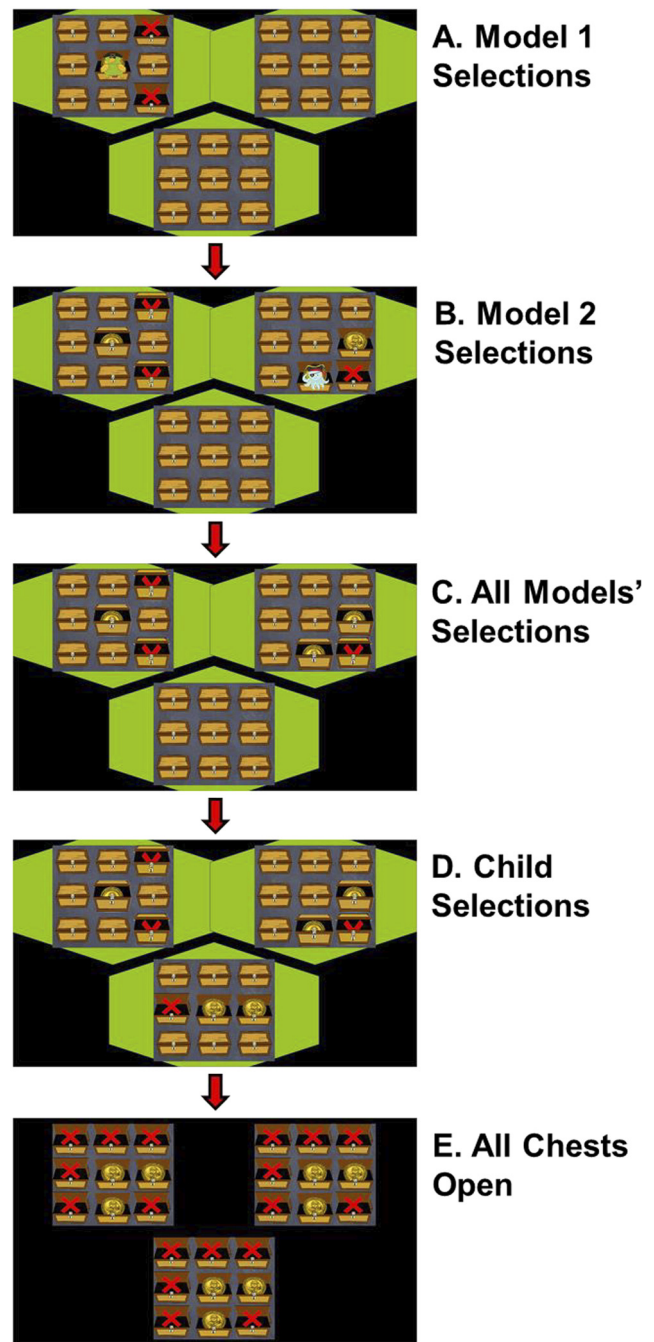


Fig. 3. Diagram of testing procedure: Model 1 selects three chests (A), Model 2 selects three chests (B), screen display following all models' selections (C), child selects three chests (D) and all the chests open to reveal the contents (E)

Note. This diagram depicts the Spatial condition as an example. The exact images displayed on screen varied depending on the condition e.g. for the Temporal and Temporal-Spatial conditions all chests in image C would be closed.

(3 per array × 24 trials, the rewards from practice trials were not included).

Ethics. This research was approved by the University of Stirling, General University Ethics Panel (reference: GUEP599). 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.

RESULTS

Firstly, we were interested in how overall use of the information provided by the two models (rewarded and unrewarded) differed with age according to the constraints presented by our four conditions (see our analysis of score and repeating). Secondly, we were interested in whether children aged 5–10 were able to make use of the information available from the two models across our four conditions to the extent that they outperformed the highest scoring single model (see “Potential for Cumulative Culture”). This question was of particular interest as a result of our motivation to understand the potential for cumulative culture. In this context, outperforming the highest scoring model could be likened to producing an improved trait variant, which goes beyond the achievements of the previous “generation”.

P-values < 0.05 were taken as statistically significant across all analyses. All generalised linear mixed effects models (GLMM) were carried out with the log link (count data, family = poisson) and the lme4 package (Bates, Mächler, Bolker, & Walker, 2015), glmer function, using R (R Core Team, 2018) version 4.1.0. Our default choice for the random effects structure for each model included by-participant random slopes for variables which varied within participant, following Barr, Levy, Scheepers, and Tily (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.

Score

We measured overall information use across our four conditions and according to child score (out of 3) per array and age (thousands of days, centred). Children were given a point for each piece of treasure found on an experimental trial, therefore the total score on each array was 0, 1, 2 or 3. This measure was useful in order to provide an overview of any effects of age and condition on overall child performance. However, it was a rather crude measure because it did not capture how children responded to demonstrations of different reward value, and to rewarded and unrewarded model selections. These were captured in our other analyses.

Score – effects of age and condition. We performed a GLMM with score (out of 3) as the dependent variable; and age and condition, and the interaction of these variables, as fixed effects. Condition was dummy coded, with each condition compared to the reference, Increased Information, condition. We also included a random intercept for the total number of rewards revealed in a trial (1, 2 or 3 across the

two models). There was a main effect of condition: lower scores (out of 3) in the Spatial ($b = -0.077$, $SE = 0.031$, $Z = -2.505$, $P < 0.050$) and the Temporal-Spatial ($b = -0.164$, $SE = 0.032$, $Z = -5.200$, $P < 0.001$) conditions compared to the reference, Increased Information, condition (Fig. 4). In the Temporal condition, the score was not significantly different from that in the reference condition ($b = -0.032$, $SE = 0.030$, $Z = -1.047$, $P = 0.295$). There was no main effect of age ($b = 0.043$, $SE = 0.035$, $Z = 1.233$, $P = 0.218$) but there was a significant interaction effect of age on the Temporal-Spatial ($b = 0.154$, $SE = 0.052$, $Z = 2.983$, $P < 0.010$) condition, showing that age had a larger affect on score in this condition, compared to the reference condition. There was no such interaction effect for the Temporal condition ($b = 0.036$, $SE = 0.050$, $Z = 0.726$, $P = 0.468$) but this was approaching significance for the Spatial condition ($b = 0.098$, $SE = 0.050$, $Z = 1.942$, $P = 0.052$). Post hoc comparisons between all four conditions were carried out using the EMMEANS package in R (Lenth, 2021). These revealed that score was significantly different for the Temporal-Spatial condition compared to the Temporal ($SE = 0.036$, $P < 0.001$) and Spatial ($SE = 0.035$, $P < 0.010$) conditions in addition to the Increased Information condition, as reported above. There was also a significant difference between the Spatial and Increased Information condition, again as reported above. There were no other significant differences between the conditions (all $P > 0.100$).

Score – effects of age by condition. We ran four exploratory GLMMs (one for each condition), with a view to further investigating the differing effects of age on score in each of the four conditions (i.e. the interaction between age and condition). We split the data by condition and removed this from the original model. The current models were therefore identical to the first model except that they had a fixed effect of age only. We found a main effect of age in the Temporal ($b = 0.079$, $SE = 0.036$, $Z = 2.233$, $P = 0.026$), Spatial ($b = 0.141$, $SE = 0.036$, $Z = 3.882$, $P < 0.001$) and Temporal-Spatial ($b = 0.197$, $SE = 0.038$, $Z = 5.193$, $P < 0.001$)

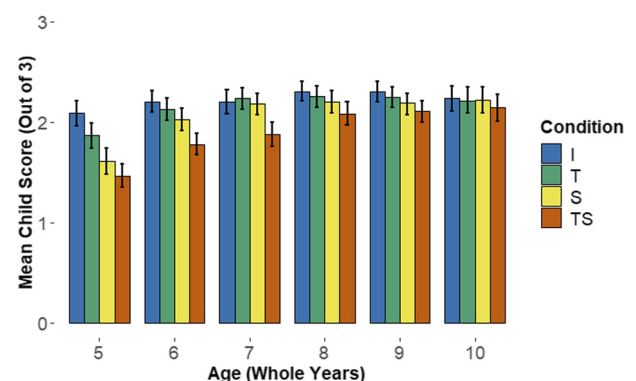
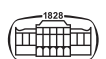


Fig. 4. Mean child Score/3 by condition and age (whole years) Note. The conditions are abbreviated as follows: I = Increased Information, T = Temporal, S = Spatial and TS = Temporal-Spatial. Error bars are 95% confidence intervals. $N = 28$ (age 5), 31 (age 6), 26 (age 7), 28 (age 8), 31 (age 9) and 20 (age 10).



conditions but not in the Increased Information condition ($b = 0.043$, $SE = 0.035$, $Z = 1.233$, $P = 0.218$). This showed that score increased with increasing age in all conditions except the Increased Information (Fig. 4).

Score – effects of condition by age. We also ran six more exploratory GLMMs (one for each age group), with a view to further investigating the differing effects of condition on score by age. We split the data by age and removed age from the original model therefore the current models were identical to the first model except that they had a fixed effect of condition only. We found main effects of conditions Temporal-Spatial ($b = -0.355$, $SE = 0.083$, $Z = -4.277$, $P < 0.001$) and Spatial ($b = -0.262$, $SE = 0.081$, $Z = -3.243$, $P = 0.001$) in our model for children aged 5: lower score in these conditions compared to the Increased Information condition (Fig. 4). For children aged 6 ($b = -0.217$, $SE = 0.074$, $Z = -2.936$, $P = 0.003$) and 7 ($b = -0.160$, $SE = 0.079$, $Z = -2.020$, $P = 0.043$) there was a main effect of the Temporal-Spatial condition only: lower score in this condition compared to the Increased Information. There were no main effects for the remaining conditions in children aged 5, 6 and 7; for 8–10-year-olds there were no main effects for any condition (all $P > 0.100$): score did not significantly differ from that in the Increased Information condition.

Repeating

Repeating the models' rewarded selections, and not repeating the models' unrewarded selections, are both correct strategies but pose quite 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 from the two models was being integrated effectively, we would expect high and low levels of repeating following rewarded and unrewarded model selections respectively. We therefore analysed the total number of rewarded and unrewarded selections repeated ("rewarded repeats" and "unrewarded repeats") on a trial by trial basis, for each participant, and how this differed according to age (thousands of days, centred) and condition (Increased Information, Temporal, Spatial, and Temporal-Spatial) in the below GLMMs.

Rewarded repeats – effects of age and condition. We performed a GLMM with the number of repeats of rewarded selections as the dependent variable; and age and condition, and the interaction of these variables, as fixed effects. Condition was dummy coded, with each condition compared to the reference, Increased Information, condition. We also included a random intercept for the total number of rewards revealed in a trial (1, 2 or 3 across the two models). There was a main effect of condition ($b = -0.154$, $SE = 0.034$, $Z = -4.498$, $P < 0.001$): fewer repeats of rewarded selections in the Temporal-Spatial condition compared to the reference, Increased Information, condition (Fig. 5A, Table 2). The Spatial and Temporal conditions were not significantly different from the reference condition (all $P > 0.100$). There

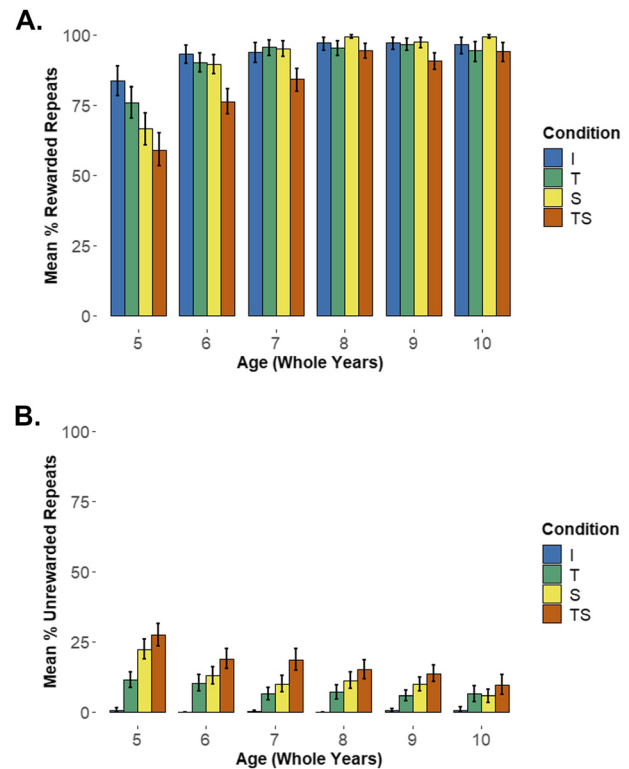


Fig. 5. Mean percentage of rewarded (Panel A) and unrewarded (Panel B) model responses which were repeated by condition and age (whole years)

Note. The conditions are abbreviated as follows: I = Increased Information, T = Temporal, S = Spatial and TS = Temporal-Spatial. Error bars are 95% confidence intervals.

was no main effect of age ($b = 0.065$, $SE = 0.038$, $Z = 1.718$, $P = 0.086$) but there was a significant interaction effect of age on the Temporal-Spatial ($b = 0.166$, $SE = 0.056$, $Z = 2.965$, $P < 0.001$) and Spatial ($b = 0.108$, $SE = 0.054$, $Z = 1.983$, $P = 0.047$) conditions, showing that age had a larger effect on the number of repeats in these conditions, compared to the reference condition. There was no such interaction effect for the Temporal condition ($b = 0.034$, $SE = 0.054$, $Z = 0.622$, $P = 0.534$). These interaction effects mirrored those of age on condition in the analysis of score (except that here the interaction between age and the difference between the Spatial and reference conditions reached significance). Therefore, to avoid repetition, we do not report further models to investigate these interaction effects here (as we did for score), but instead in the SI. Post hoc comparisons between all four conditions were carried out using the EMMEANS package in R (Lenth, 2021). These revealed that the number of rewarded repeats was significantly different for the Temporal-Spatial condition compared to the Temporal ($SE = 0.031$, $P < 0.001$) and Spatial ($SE = 0.031$, $P = 0.001$) conditions in addition to the Increased Information condition, as reported above. There were no other significant differences between the conditions (all $P > 0.100$).

Unrewarded repeats – effects of age and condition. We performed a GLMM which was identical to the model in section "rewarded repeats – effects of age and condition"

Table 2. Mean Percentage of Rewarded and Unrewarded Model Selections Repeated (and Standard Deviation) for Ages 5–10 (Whole Years) in the Increased Information (I), Temporal (T), Spatial (S) and Temporal-Spatial (TS) Conditions

Age (whole years)	Condition	Mean percentage rewarded repeats	SD (rewarded repeats)	Mean percentage unrewarded repeats	SD (unrewarded repeats)
5	I	83.730	35.646	0.595	5.439
6	I	93.100	22.698	0.000	0.000
7	I	93.910	22.085	0.214	2.669
8	I	97.123	15.900	0.000	0.000
9	I	97.222	14.798	0.448	4.395
10	I	96.667	18.026	0.556	6.086
5	T	75.893	36.147	11.310	19.671
6	T	90.233	24.906	10.215	20.582
7	T	95.727	17.948	6.410	14.475
8	T	95.437	17.656	6.944	15.414
9	T	96.595	15.608	5.824	14.410
10	T	94.444	19.842	6.389	16.481
5	S	66.567	39.436	22.222	23.758
6	S	89.427	24.481	12.903	21.029
7	S	95.085	17.982	9.829	17.747
8	S	99.405	4.428	11.111	19.785
9	S	97.491	13.071	9.677	18.063
10	S	99.444	4.285	5.694	14.085
5	TS	58.929	38.986	27.480	27.306
6	TS	76.254	32.638	18.907	24.188
7	TS	84.295	26.790	18.483	23.915
8	TS	94.444	17.933	14.980	21.783
9	TS	90.681	21.289	13.620	20.931
10	TS	94.167	19.404	9.583	18.418

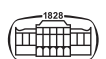
except that we used the number of unrewarded repeats as the dependent variable. There was a main effect of condition: more unrewarded repeats in the Temporal ($b = 3.397$, $SE = 0.396$, $Z = 8.575$, $P < 0.001$), Spatial ($b = 3.809$, $SE = 0.394$, $Z = 9.665$, $P < 0.001$) and Temporal-Spatial ($b = 4.170$, $SE = 0.393$, $Z = 10.618$, $P < 0.001$) conditions compared to the reference, Increased Information, condition (Fig. 5B, Table 2). There was no main effect of age ($b = 0.414$, $SE = 0.628$, $Z = 0.659$, $P = 0.510$) and no interaction effects between age and the Temporal, Spatial or Temporal-Spatial conditions (all $P > 0.100$). Post hoc comparisons between all four conditions were carried out using the EMMEANS package in R (Lenth, 2021). These revealed that the number of unrewarded repeats was significantly greater for each of the conditions compared to the Increased Information condition, as reported above. Each condition also differed significantly from each of the other conditions: there were more unrewarded repeats in the Spatial condition than the Temporal ($SE = 0.061$, $P < 0.001$) and in the Temporal-

Spatial condition compared to both the Spatial ($SE = 0.053$, $P < 0.001$) and Temporal conditions ($SE = 0.040$, $P < 0.001$).

Potential for cumulative culture

In order to measure children's potential for cumulative culture, based on the different patterns of performance outlined in our introduction, we measured whether the mean child score (out of 3) on the model-score-2 trials (for each age group and condition) adhered to the criteria outlined in four levels. Mean score needed to be as follows to achieve each level: Level 0: at or below the chance-level of 1; Level 1: significantly greater than chance-level; Level 2: significantly greater than both chance-level and the mean score across the two models; Level 3: significantly greater than the highest single model score.

For the model-score-2 trials (Figs 6 and 7; Table S1) children performed most proficiently in the Increased Information (control) condition in which all age groups scored



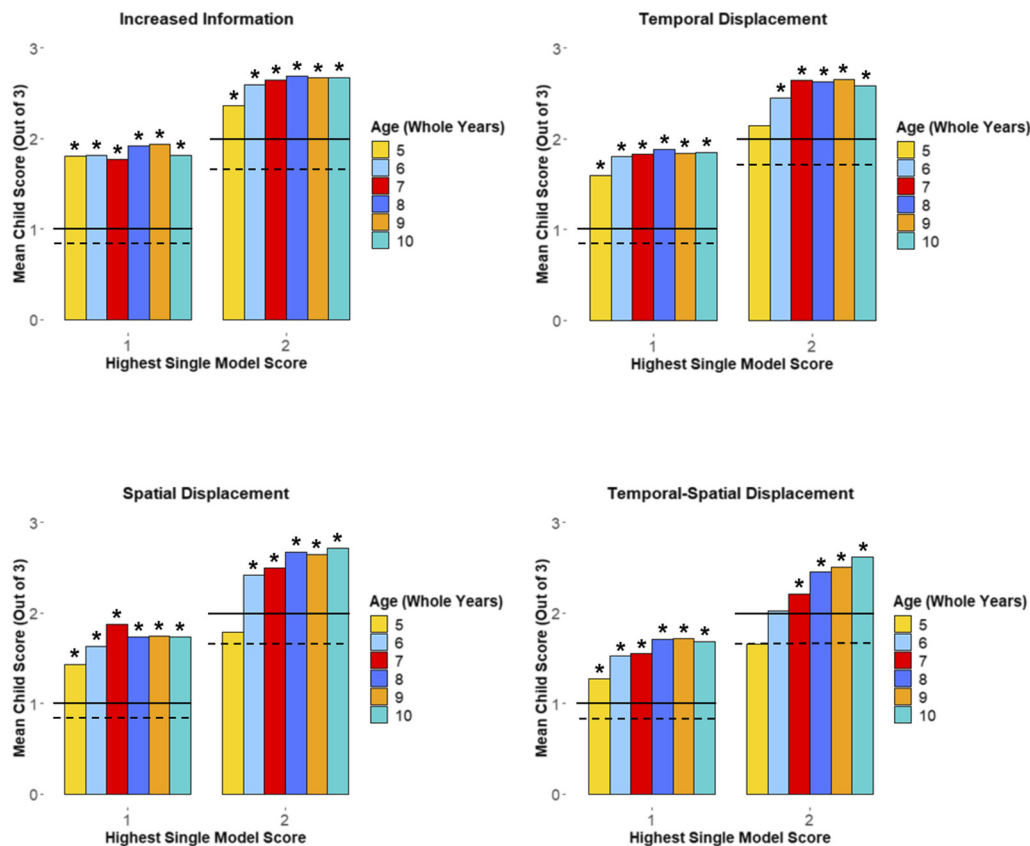


Fig. 6. Mean child Score/3 by age (whole years) and the highest single model score in a trial for the Increased Information, Temporal, Spatial and Temporal-Spatial conditions

Note. The solid line at 1 depicts chance performance: the score expected if children were selecting chests at random and not using the social information. The dashed lines at 0.83 (model-score-1 trials) and 1.67 (model-score-2 trials) depict the mean scores across the two models in the respective trials. The solid line at 2 allows visualisation of whether children scored above 2 when the highest single model score in a trial was 2. An asterisk indicates that the mean child score/3 is significantly above 1 or 2 for the model-score-1 and model-score-2 trials respectively.

significantly higher than the highest single model score of 2, hence achieved Level 3 (ages 6–10, $P < 0.001$; age 5, $P = 0.011$; Table S1). However, in the Temporal condition the level achieved differed with age. Ages 6–10 achieved Level 3 (all P values < 0.001 ; Table S1). Children aged 5 did not outperform the highest single model score ($P = 0.145$; Table S1) but achieved a Level 2 because they performed significantly above chance level ($P < 0.001$) and significantly above the mean score across both models ($P < 0.001$). This showed that they were performing better than naïve exploration and gaining some performance benefit from the presence of two models. In the Spatial condition ages 6–10 again achieved a Level 3 (all P values < 0.001 ; Table S1). Children aged 5 did not outperform the highest single model score ($P = 0.941$; Table S1) and although they performed significantly above chance ($P < 0.001$) they did not significantly outperform the mean score across the models ($P = 0.190$) and were therefore given a Level 1. In the Temporal-Spatial condition 5-year-olds again scored a Level 1 i.e. they did not outperform the highest single model score ($P = 0.941$; Table S1) or the mean score across the models ($P = 0.547$) but they did outperform chance ($P < 0.001$).

However, 6-year-olds performed more poorly than in all other conditions, achieving a Level 2 because they did not outperform the highest single model score ($P = 0.411$; Table S1) but outperformed both the mean score ($P < 0.001$) and chance ($P < 0.001$). Outperformance of the highest single model score, Level 3, was therefore found only in ages 7–10 (ages 8–10, $P < 0.001$; age 7, $P = 0.010$; Table S1).

As stated above, for the model-score-1 trials it was only useful to determine if children scored more highly than chance (1.0) and therefore whether they were gaining some benefit from the information provided by the models (equivalent to Level 1). As in the model-score-2 trials, each age group, in each condition, performed significantly above chance-level; ages 6–10, $P < 0.001$; age 5, $P = 0.004$ (Fig. 6, Table S2).

Across the six demonstration types (outlined in section “demonstration types”), three were model-score-1 trials (3, 5 and 6, Table 1), and three model-score-2 trials (1, 2 and 4, Table 1). The three demonstration types present within these model-score-1/model-score-2 groups were obviously identical in terms of highest single model score but differed

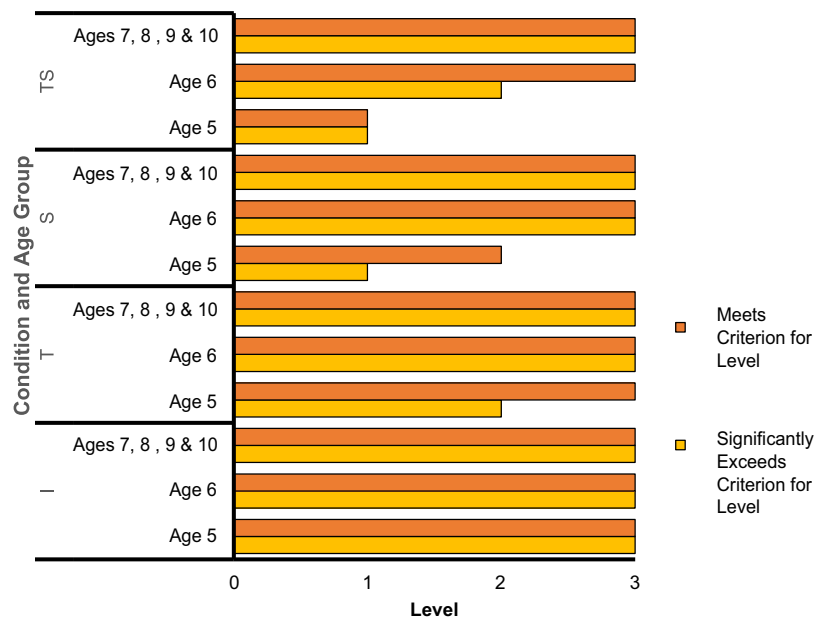


Fig. 7. Potential for cumulative culture performance classification level by condition and age group

Note. Situations in which a mean child score/3 was *significantly* higher (yellow bars) and *higher* but not *significantly* higher (orange 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.

in factors such as the total number of rewarded locations revealed and the number of rewarded/unrewarded selections repeated etc. (see Table 1 for specific differences). In addition to the above analysis for the model-score-1 and model-score-2 groups, we therefore plotted how mean score differed according to the six different demonstration types (Figure S1), and according to the total number of rewarded model selections in a trial – 1, 2 or 3 (Figure S2). Breaking down the results in these additional ways did not reveal any patterns within the data which were not already captured in our other analyses. As in the analyses already reported, effects of age were apparent in the Temporal, Spatial and Temporal-Spatial conditions, particularly for higher scoring demonstrations.

DISCUSSION

Capitalising on increases in population size by utilising information from multiple models may be an important mechanism by which improved traits are introduced into human populations. We reasoned that exploiting this increase in information may require human-unique cognitive capacities (e.g. enhanced use of working memory, metacognition or mental translation) because a larger cognitive load is likely generated when information from more than one source is separated in time and/or space. We thus presented children with Temporal, Spatial and Temporal-Spatial displacement conditions which emulated real-life constraints on information use and undoubtedly increased the cognitive load associated with obtaining maximum benefit from information provided by multiple models. Our

novel method allowed investigation of the ability to utilise this displaced information from two models, and moreover, children’s potential to outperform the best available information within this multiple-model population. We predicted that this would be possible across a wider range of conditions as age, and thus cognitive capacity, increased. Such a result would suggest that enhanced cognitive abilities (such as those outlined above) may better enable humans to exploit information provided by multiple models, despite the associated challenges, to generate new trait variants and drive cumulative culture.

Our results were broadly consistent with the above predictions. Firstly, in our least cognitively challenging control (Increased Information) condition, children utilised rewarded and unrewarded information at high levels right across our age range (5–10 years) and score did not differ according to age. Moreover, all ages evidenced outperformance of the highest single scoring model (i.e. the best available information in the population) and achieved a Level 3 on our continuum of potential for cumulative culture. Children therefore displayed the ability to utilise information from the two models to the extent that they were able to generate an improved trait variant (a higher score) when the information was not displaced in time or space (i.e. when the information from both models remained present for use within the same search space, presumably creating only minimal cognitive load). In line with results by Subiaul et al. (2015), this demonstrated that even the youngest children did not find using information from two models difficult per se – rather, any difficulty in making the most advantageous use of information from multiple models may lie in the extra cognitive resources required to store, retrieve, and spatially

translate information. In real-life social learning scenarios, we would expect the information from multiple models to be displaced in time and/or space therefore, although this result confirms that children can use information from two models in theory, it probably reveals little regarding children's propensity to do so in real-life.

Secondly, also in line with our predictions, we found that score and repeating of rewarded selections increased with age in the conditions which presented a greater cognitive load (Spatial, Temporal and Temporal-Spatial). The youngest children's performance was particularly weak in the Spatial and Temporal-Spatial conditions: 5-year-olds were the only age group with a lower score in both of these conditions compared to the control condition, and 6- and 7-year-olds scored more poorly in the Temporal-Spatial condition. However, older children's (8–10 years) score did not differ from the control in any condition. Additionally, older children (7–10 years) outperformed the highest scoring model and thus showed evidence of cumulative improvement across a wider range of conditions than younger children – in the Temporal, Spatial and Temporal-Spatial conditions (Level 3 on our continuum of potential for cumulative culture). Although 6-year-olds outperformed 5-year-olds, their performance did not match that of 7–10-year-olds in the Temporal-Spatial condition – they repeated significantly fewer rewarded selections in comparison to the Increased Information condition and achieved Level 2. 6-year-olds therefore showed evidence consistent with the potential for cumulative culture in the Increased Information, Temporal and Spatial conditions only. However, 5-year-olds performed the most poorly – they appeared to gain little benefit from the presence of multiple models (achieving only Level 1) in either the Spatial or Temporal-Spatial conditions. They performed better in the Temporal condition (achieving Level 2) but still not to the extent that an improved trait variant was created.

As far as we are aware, our study is the first to demonstrate that children's ability to utilise information from multiple models may be affected by the types of cognitive constraints which exist in real-life social learning scenarios. Moreover, we are the first to show that children's ability to outperform the highest scoring single model in a multiple-model population, showing the potential for cumulative culture, changes with development (age 5–10) and the level of cognitive load (task constraints) presented. This is consistent with our previous work (Wilks et al., 2021) in which we found that children's capacity to use the social information provided by one model depended both on age and memory constraints (which differed across two task conditions). We thus postulate that, in any given experimental task or real-life scenario, the constraints on accessing information (e.g. temporal or spatial separation), and cognitive ability, will determine whether these constraints can be overcome, allowing the information to be used and cumulative improvements to occur. We would expect that the ability to overcome particular constraints will differ depending on the cognitive abilities of the population, e.g. adults, children at different stages of development, or nonhumans. In the case of accessing

information from multiple models, this hypothesis is also supported by the adult experimental literature. Kempe and Mesoudi (2014) found that adults could integrate spatially separated information, leading to cumulative culture. Yet studies by Caldwell and Millen (2010) and Fay et al. (2019) highlighted that there may be constraints on adults' ability to integrate information if they do not have sufficient time in which to process it or the memory load is too great.

So why was the performance of 5-year-olds in our displacement conditions (and 6-year-olds, in the Temporal-Spatial condition only) poor in comparison to that of the other age groups? We postulate that the increased cognitive load presented by the temporal, spatial and (to a greater extent) temporal-spatial separation of the models' selections was difficult to overcome at this stage in a child's cognitive development. In our Temporal displacement conditions, age-related improvements to working memory (Best & Miller, 2010; Diamond, 2013; Garon et al., 2008) may have better enabled older children to hold the information provided by both models in mind whilst deciding how best to utilise it. Furthermore, the continued development of a range of metacognitive processes and strategies (e.g. mnemonic devices (Jurowski et al., 2015), or inner speech (Carruthers, 2013; Cowan, 2008), see introduction) may have effectively reduced the cognitive load through "cognitive offloading" (Armitage et al., 2020; Bulley et al., 2020). Increased task success from age 6 (although note that 6-year-olds did struggle to make full use of the available information in the Temporal-Spatial condition with the highest cognitive load) may have resulted from metacognitive developments in understanding that access to partial perceptual information does not lead to complete knowledge (Chandler & Helm, 1984; Kloo et al., 2017; Rohwer et al., 2012; Sodian & Wimmer, 1987; Taylor, 1988). When utilising information from two models, this insight could have encouraged children to be more selective in their use of information from each model.

We would expect the performance of nonhumans to be similar to, or poorer than, that of 5-year-olds due to limitations in accessing and using the displaced information. Firstly, with respect to working memory, it is difficult to directly compare humans and nonhumans due to the small amount of comparative research available for analysis. However, it appears that, although some nonhuman mammals may have similar storage capacities, humans can represent concepts within memory differently (e.g. through the use of mnemonic devices, Jurowski et al., 2015, or inner speech, Carruthers, 2013; Cowan, 2008) and have a better ability to deploy attention and resist interference (for a systematic analysis of evidence to date see Carruthers, 2013). It is therefore likely that when presented with temporally displaced information nonhuman primates would perform similarly to young children, who are still developing such capacities, and would not be expected to produce an improved variant under temporal displacement.

Secondly, regarding metacognition, there is growing evidence from naturalistic information-seeking (Call & Carpenter, 2001) and uncertainty monitoring (Smith, Shields, Schull, & Washburn, 1997) paradigms, that some



nonhumans (e.g. chimpanzees, orangutans, Bohn, Allritz, Call, & Völter, 2017; Call & Carpenter, 2001, gorillas, bonobos, Call, 2010, and rhesus monkeys, Beran, Smith, Redford, & Washburn, 2006; Couchman, Coutinho, Beran, & Smith, 2010; Hampton, Zivin, & Murray, 2004) may have the ability to accurately monitor their own uncertainty and make simple responses according to this (Beran, Decker, Schwartz, & Smith, 2012). However, these tasks require more limited information processing than that we have presented – the rewards are not split across multiple locations and the required responses are simpler. The ability to make such information-seeking and uncertainty monitoring responses is unlikely to allow for devising and using metacognitive strategies as may increase information use in this task, e.g. the aforementioned memory aids such as mnemonic devices or inner speech. Furthermore, the representational nature of this uncertainty is unknown (Beran et al., 2012; Carruthers, 2008, 2009) and (considering findings regarding animal mindreading, Call & Tomasello, 2008; Heyes, 2015) it seems unlikely that nonhumans explicitly represent “I know” (or a non-linguistic equivalent of such a belief) as humans do. This would be expected to limit their ability to make appropriately selective responses in more subtle situations (e.g. that presented in this study) in which partial information is available from multiple sources. Nevertheless, we might expect some nonhumans to be able to utilise information from multiple models under more limited circumstances, such as when the information is not displaced (e.g. in our Increased Information condition or that presented in Subiaul et al., 2015). This demands further research.

As noted in our introduction, making maximal use of multiple pieces of information separated in space likely requires one to translate the observed information to one’s own bodily frame of reference. This would be expected to increase the associated cognitive load due to requirements such as mental translation (Frick et al., 2013; Levine et al., 1999; Shepard & Metzler, 1971) and the ability to understand dual representation (DeLoache, 1989, 1991, 2000). In our Spatial and Temporal-Spatial conditions these skills are likely to be required for understanding that the models’ selections relate to each other and to comparative locations in one’s own search space. Moreover, we expect these will be needed for accurately translating relevant information (rewarded selections) from a model’s search space to one’s own. Research has shown that mentally moving visual information is cognitively taxing – e.g. there is a positive, linear relationship between the time taken to mentally rotate a shape and the angle through which it must be rotated (Shepard & Metzler, 1971). Such ability is considered to be present from age 5 (Frick et al., 2013; Iachini, Ruggiero, Bartolo, Rapuano, & Ruotolo, 2019; Marmor, 1975; evidence in younger children is inconsistent, Estes, 1998; Frick, Daum, Walser, & Mast, 2009, 2013) and to continue to develop throughout childhood and into adolescence (Kail, Pellegrino, & Carter, 1980). This developmental trajectory thus supports our finding that utilising the spatially separated information was challenging for the youngest children, especially when combined with the added constraint of

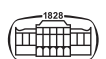
temporal separation. We would therefore expect nonhuman primates to perform similarly due to the increased cognitive load required to translate information from multiple locations separated in space. However, further research would be needed to ascertain this as there is some evidence that nonhumans can engage in mental rotation (Köhler, Hoffmann, Dehnhardt, & Mauck, 2005; Stich, Dehnhardt, & Mauck, 2003; Vauclair, Fagot, & Hopkins, 1993), although currently no evidence that this is homologous to the process in humans (Carruthers, 2013).

We have shown that children aged 6–10-years-old can utilise information from multiple models to generate improved variants under ecologically valid constraints. However, it is unlikely that all new cultural traits are derived by combining information across different models, and some traits may not lend themselves to such combinatory mechanisms. To reiterate, we do not argue that the kind of combinatory social learning investigated in this study is the only method by which improved variants may arise within populations, but rather, that human cognition may enable our species to exploit information from multiple models and thus engage in this method.

We used identifiably different demonstrators (a cartoon parrot and octopus) for the two demonstrations on each trial of our task because wanted the multiple models feature to be as transparent and concrete as possible. Yet, it remains possible that we would obtain a different result were the information presented to children in the absence of demonstrators; that is, the demonstrations occurred without a social element. Future work could test this using our paradigm. However, we would not necessarily predict a different result provided the children were still motivated to use the information in the demonstrations, and the information itself remained the same. In this sense, the demonstrators could be considered to act primarily to set the information in context and to reinforce the task goal and the desirability of the reward. In support of this prediction, previous work from our group (Atkinson et al., 2020; Renner, Kean, Atkinson, & Caldwell, 2021) found that children’s use of information in a logically similar task was unaffected by its source.

Theories which posit that population size underlies human cumulative culture (Henrich, 2004; Powell et al., 2009; Shennan, 2001) do not directly consider the underlying cognitive abilities needed to exploit information from multiple social models. However, we have shown that deriving full benefit from the increased information provided by multiple models may be dependent on the ability to store, manipulate and retrieve this information. It therefore follows that the effects of population size on cumulative culture may be inherently dependent on the cognitive abilities of the learners.

Ethics statement: This research was approved by the University of Stirling, General University Ethics Panel (reference: GUEP599). 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.



Declarations of interest: None.

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 data collection: Rhea Paterson.

SUPPLEMENTARY MATERIAL

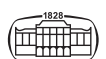
Supplementary data to this article can be found online at: <https://doi.org/10.1556/2055.2021.00005>.

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*. <https://doi.org/10.1037/xge0000959>.
- Barr, D., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 1–43. <https://doi.org/10.1016/j.jml.2012.11.001>.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Beran, M. J., Decker, S., Schwartz, A., & Smith, J. D. (2012). Uncertainty monitoring by young children in a computerized task. *Scientifica*, 2012, 1–6. <https://doi.org/10.6064/2012/692890>.
- Beran, M. J., Smith, J. D., Redford, J. S., & Washburn, D. A. (2006). Rhesus macaques (*Macaca mulatta*) monitor uncertainty during numerosity judgments. *Journal of Experimental Psychology. Animal Behavior Processes*, 32(2), 111–119. <https://doi.org/10.1037/0097-7403.32.2.111>.
- 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>.
- Blakey, K.H., Rafetseder, E., Atkinson, M., Renner, E., Cowan-Forsythe, F., Sati, S.J., & Caldwell, C.A. (2020). Development of Strategic Social Information Seeking: Implications for Cumulative Culture. *PsyArXiv*. <https://doi.org/10.31234/osf.io/4envx>.
- Bohn, M., Allritz, M., Call, J., & Völter, C. J. (2017). Information seeking about tool properties in great apes. *Scientific Reports*, 7(1), 10923. <https://doi.org/10.1038/s41598-017-11400-z>.
- Bryce, D., & Whitebread, D. (2012). The development of meta-cognitive skills: Evidence from observational analysis of young children's behavior during problem-solving. *Metacognition and Learning*, 7(3), 197–217. <https://doi.org/10.1007/s11409-012-9091-2>.
- Buchsbaum, D., Gopnik, A., Griffiths, T. L., & Shafto, P. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition*, 120(3), 331–340. <https://doi.org/10.1016/j.cognition.2010.12.001>.
- 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., & Millen, A. E. (2010). Human cumulative culture in the laboratory: Effects of (micro) population size. *Learning & Behavior: A Psychonomic Society Publication*, 38(3), 310–318. <https://doi.org/10.3758/LB.38.3.310>.
- Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen? *Animal Cognition*, 3(4), 207–220. <https://doi.org/10.1007/s100710100078>.
- Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, 12(5), 187–192. <https://doi.org/10.1016/j.tics.2008.02.010>.
- Carruthers, P. (2008). Meta-cognition in animals: A skeptical look. *Mind and Language*, 23(1), 58–89. <https://doi.org/10.1111/j.1468-0017.2007.00329.x>.
- Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, 32(2), 121–138. <https://doi.org/10.1017/S0140525X09000545>.
- Carruthers, P. (2013). Evolution of working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 110(suppl. 2), 10371–10378. <https://doi.org/10.1073/pnas.1301195110>.
- Chandler, M. J., & Helm, D. (1984). Developmental changes in the contribution of shared experience to social role-taking competence. *International Journal of Behavioral Development*, 7(2), 145–156. <https://doi.org/10.1177/016502548400700203>.
- Claidière, N., Smith, K., Kirby, S., & Fagot, J. (2014). Cultural evolution of systematically structured behaviour in a non-human primate. *Proceedings of the Royal Society B*, 281, 20141541–20141541. <https://doi.org/10.1098/rspb.2014.1541>.
- Collard, M., Ruttle, A., Buchanan, B., & O'Brien, M. J. (2013). Population size and cultural evolution in nonindustrial food-producing societies. *Plos One*, 8(9), 1–6. <https://doi.org/10.1371/journal.pone.0072628>.
- Collard, M., Vaesen, K., Cosgrove, R., & Roebroeks, W. (2016). The empirical case against the 'demographic turn' in palaeolithic archaeology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1698). <https://doi.org/10.1098/rstb.2015.0242>.
- Couchman, J. J., Coutinho, M. V. C., Beran, M. J., & Smith, J. D. (2010). Beyond stimulus cues and reinforcement signals: A new approach to animal metacognition. *Journal of Comparative Psychology*, 124(4), 356–368. <https://doi.org/10.1037/a0020129>.
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? Nelson. *NIH Public Access*,



- 6123(07), 323–338. [https://doi.org/10.1016/S0079-6123\(07\)00020-9](https://doi.org/10.1016/S0079-6123(07)00020-9).
- Dean, L. G., Vale, G. L., Laland, K. N., Flynn, E., & Kendal, R. L. (2014). Human cumulative culture: A comparative perspective. *Biological Reviews*, 89(2), 284–301. <https://doi.org/10.1111/brv.12053>.
- DeLoache, J. S. (1989). Young children's understanding of the correspondence between a scale model and a larger space. *Cognitive Development*, 4(2), 121–139. [https://doi.org/10.1016/0885-2014\(89\)90012-9](https://doi.org/10.1016/0885-2014(89)90012-9).
- DeLoache, J. S. (1991). Symbolic functioning in very young children: Understanding of pictures and models. *Child Development*, 62(4), 736–752.
- DeLoache, J. S. (2000). Dual representation and young children's use of scale models. *Child Development*, 71(2), 329–338. <https://doi.org/10.1111/1467-8624.00148>.
- Derex, M., Beugin, M. P., Godelle, B., & Raymond, M. (2013). Experimental evidence for the influence of group size on cultural complexity. *Nature*, 503(7476), 389–391. <https://doi.org/10.1038/nature12774>.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64(1), 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>.
- Estes, D. (1998). Young children's awareness of their mental activity: The case of mental rotation. *Child Development*, 69(5), 1345–1360. <https://doi.org/10.1111/j.1467-8624.1998.tb06216.x>.
- Fay, N., De Kleine, N., Walker, B., & Caldwell, C. A. (2019). Increasing population size can inhibit cumulative cultural evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 116(14), 6726–6731. <https://doi.org/10.1073/pnas.1811413116>.
- Frick, A., Daum, M. M., Walser, S., & Mast, F. W. (2009). Motor processes in children's mental rotation. *Journal of Cognition and Development*, 10(1–2), 18–40. <https://doi.org/10.1080/15248370902966719>.
- Frick, A., Hansen, M. A., & Newcombe, N. S. (2013). Development of mental rotation in 3- to 5-year-old children. *Cognitive Development*, 28(4), 386–399. <https://doi.org/10.1016/j.cogdev.2013.06.002>.
- 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>.
- Hampton, R. R., Zivin, A., & Murray, E. A. (2004). Rhesus monkeys (*Macaca mulatta*) discriminate between knowing and not knowing and collect information as needed before acting. *Animal Cognition*, 7(4), 239–246. <https://doi.org/10.1007/s10071-004-0215-1>.
- Henrich, J. (2004). Demography and cultural evolution: How adaptive cultural processes can produce maladaptive losses—the Tasmanian case. *American Antiquity*, 69(2), 197–214. <https://doi.org/10.2307/4128416>.
- Henrich, J., Boyd, R., Derex, M., Kline, M. A., Mesoudi, A., Muthukrishna, M., ... Thomas, M. G. (2016). Understanding cumulative cultural evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 113(44), E6724–E6725. <https://doi.org/10.1073/pnas.1610005113>.
- Heyes, C. (2015). Animal mindreading: What's the problem? *Psychonomic Bulletin & Review*, 22(2), 313–327. <https://doi.org/10.3758/s13423-014-0704-4>.
- Iachini, T., Ruggiero, G., Bartolo, A., Rapuano, M., & Ruotolo, F. (2019). The effect of body-related stimuli on mental rotation in children, young and elderly adults. *Scientific Reports*, 9(1), 3–12. <https://doi.org/10.1038/s41598-018-37729-7>.
- Jesmer, B. R., Merkle, J. A., Goheen, J. R., Aikens, E. O., Beck, J. L., Courtemanch, A. B., ... Kauffman, M. J. (2018). Is ungulate migration culturally transmitted? Evidence of social learning from translocated animals. *Science*, 361(6406), 1023–1025. <https://doi.org/10.1126/science.aat0985>.
- Jurowski, K., Jurowska, A., Krzeczowska, M., Jurowski, K., Jurowska, A., & Krzeczowska, M. (2015). Comprehensive review of mnemonic devices and their applications: State of the art. *International E-Journal of Science, Medicine and Education*, 9(3), 4–9.
- Kail, R., Pellegrino, J., & Carter, P. (1980). Developmental changes in mental rotation. *Journal of Experimental Child Psychology*, 29(1), 102–116. [https://doi.org/10.1016/0022-0965\(80\)90094-6](https://doi.org/10.1016/0022-0965(80)90094-6).
- Kempe, M., & Mesoudi, A. (2014). An experimental demonstration of the effect of group size on cultural accumulation. *Evolution and Human Behavior*, 35(4), 285–290. <https://doi.org/10.1016/j.evolhumbehav.2014.02.009>.
- Kline, M. A., & Boyd, R. (2010). Population size predicts technological complexity in Oceania. *Proceedings of the Royal Society B: Biological Sciences*, 277(1693), 2559–2564. <https://doi.org/10.1098/rspb.2010.0452>.
- Kloo, D., Rohwer, M., & Perner, J. (2017). Direct and indirect admission of ignorance by children. *Journal of Experimental Child Psychology*, 159, 279–295. <https://doi.org/10.1016/j.jecp.2017.02.014>.
- Kobayashi, Y., & Aoki, K. (2012). Innovativeness, population size and cumulative cultural evolution. *Theoretical Population Biology*, 82(1), 38–47. <https://doi.org/10.1016/j.tpb.2012.04.001>.
- Köhler, C., Hoffmann, K. P., Dehnhardt, G., & Mauck, B. (2005). Mental rotation and rotational invariance in the Rhesus monkey (*Macaca mulatta*). *Brain, Behavior and Evolution*, 66(3), 158–166. <https://doi.org/10.1159/000087156>.
- Lake, M. W., & Venti, J. (2009). Quantitative analysis of macro-evolutionary patterning in technological evolution: Bicycle design from 1800 to 2000. In S. J. Shennan (Ed.), *Pattern and Process in Cultural Evolution* (pp. 147–174). University of California Press.
- Lenth, R. V. (2021). emmeans: Estimated marginal means, aka least-squares means. R package version 1.6.2-1. <https://CRAN.R-project.org/package=emmeans>.
- Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology*, 35(4), 940–949. <https://doi.org/10.1037/0012-1649.35.4.940>.
- Lewis, H. M., & Laland, K. N. (2012). Transmission fidelity is the key to the build-up of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1599), 2171–2180. <https://doi.org/10.1098/rstb.2012.0119>.
- Marmor, G. S. (1975). Development of kinetic images: When does the child first represent movement in mental images? *Cognitive Psychology*, 7, 548–559. [https://doi.org/10.1016/0010-0285\(75\)90022-5](https://doi.org/10.1016/0010-0285(75)90022-5).



- Mesoudi, A., & Thornton, A. (2018). What is cumulative cultural evolution? *Proceedings of the Royal Society B: Biological Sciences*, 285(1880). <https://doi.org/10.1098/rspb.2018.0712>.
- Muthukrishna, M., & Henrich, J. (2016). Innovation in the collective brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1690). <https://doi.org/10.1098/rstb.2015.0192>.
- Muthukrishna, M., Shulman, B. W., Vasilescu, V., & Henrich, J. (2014). Sociality influences cultural complexity. *Proceedings. Biological Sciences*, 281(1774), 20132511. <https://doi.org/10.1098/rspb.2013.2511>.
- O'Brien, M. J., & Lyman, R. L. (2000). Darwinian evolutionism is applicable to historical archaeology. *International Journal of Historical Archaeology*, 4(1), 71–112. <https://doi.org/10.1023/A:1009556427520>.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>.
- Powell, A., Shennan, S., & Thomas, M. G. (2009). Late pleistocene demography and the appearance of modern human behavior. *Science*, 324(5932), 1298–1301. <https://doi.org/10.1126/science.1170165>.
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Renner, E., Kean, D., Atkinson, M., & Caldwell, C. A. (2021). The use of individual, social, and animated cue information by capuchin monkeys and children in a touchscreen task. *Scientific Reports*, 11(1043). <https://doi.org/10.1038/s41598-020-80221-4>.
- Rohwer, M., Kloo, D., & Perner, J. (2012). Escape from meta-ignorance: How children develop an understanding of their own lack of knowledge. *Child Development*, 83(6), 1869–1883. <https://doi.org/10.1111/j.1467-8624.2012.01830.x>.
- Sasaki, T., & Biro, D. (2017). Cumulative culture can emerge from collective intelligence in animal groups. *Nature Communications*, 8, 1–6. <https://doi.org/10.1038/ncomms15049>.
- Schofield, D. P., McGrew, W. C., Takahashi, A., & Hirata, S. (2017). Cumulative culture in nonhumans: Overlooked findings from Japanese monkeys? *Primates*, 59(2), 113–122. <https://doi.org/10.1007/s10329-017-0642-7>.
- Shennan, S. (2001). Demography and cultural innovation: A model and its implications for the emergence of modern human culture. *Cambridge Archaeological Journal*, 11(1), 5–16. <https://doi.org/10.1017/s0959774301000014>.
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701–703.
- Smith, J. D., Shields, W. E., Schull, J., & Washburn, D. A. (1997). The uncertain response in humans and animals. *Cognition*, 62(1), 75–97. [https://doi.org/10.1016/S0010-0277\(96\)00726-3](https://doi.org/10.1016/S0010-0277(96)00726-3).
- Sodian, B., & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. *Child Development*, 58(2), 424–433. <https://doi.org/10.2307/1130519>.
- Stich, K. P., Dehnhardt, G., & Mauck, B. (2003). Mental rotation of perspective stimuli in a California sea lion (*Zalophus californianus*). *Brain, Behavior and Evolution*, 61(2), 102–112. <https://doi.org/10.1159/000069355>.
- Subiaul, F., Krajkowski, E., Price, E. E., & Etz, A. (2015). Imitation by combination: Preschool age children evidence summative imitation in a novel problem-solving task. *Frontiers in Psychology*, 6, 1–13. <https://doi.org/10.3389/fpsyg.2015.01410>.
- Subiaul, F., & Stanton, M. A. (2020). Intuitive invention by summative imitation in children and adults. *Cognition*, 202. <https://doi.org/10.1016/j.cognition.2020.104320>.
- Taylor, M. (1988). The development of children's understanding of the seeing-knowing distinction. In Astington, J. W., Harris, P. L., & Olson, D. R. (Eds.), *Developing theories of mind* (pp. 207–225). Cambridge University Press.
- Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: On the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1528), 2405–2415. <https://doi.org/10.1098/rstb.2009.0052>.
- 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 University Press. <https://doi.org/10.1017/CBO9780511665486.012>.
- Vaesen, K., Collard, M., Cosgrove, R., & Roebroeks, W. (2016). Population size does not explain past changes in cultural complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 113(16), E4241–E4247. <https://doi.org/10.1073/pnas.1520288113>.
- Vauclair, J., Fagot, J., & Hopkins, W. D. (1993). Rotation of mental images in baboons when the visual input is directed to the left cerebral hemisphere. *Psychological Science*, 4(2), 99–103. <https://doi.org/10.1111/j.1467-9280.1993.tb00468.x>.
- Wilks, C.E.H., Rafetseder, E., Renner, E., Atkinson, M., & Caldwell, C.A. (2021). Cognitive Prerequisites for Cumulative Culture are Context-Dependent: Children's Potential for Ratcheting Depends on Cue Longevity. *Journal of Experimental Child Psychology*, E105031. <https://doi.org/10.1016/j.jecp.2020.105031>.
- Winters, J. (2020). Is the cultural evolution of technology cumulative or combinatorial? <https://doi.org/10.31235/osf.io/aypnx>.
- Wood, L. A., Kendal, R. L., & Flynn, E. G. (2013). Whom do children copy? Model-based biases in social learning. *Developmental Review*, 33(4), 341–356. <https://doi.org/10.1016/j.dr.2013.08.002>.
- Woods, D.L., Kishiyama, M.M., Yund, E.W., Herron, T.J., Edwards, B., Poliva, O., ... Reed, B. (2011). Improving digit span assessment of short-term verbal memory. *Journal of Clinical and Experimental Neuropsychology*, 33(1), 101–111. <https://doi.org/10.1080/13803395.2010.493149>.

