The Roots of the Ratchet:

An Evaluation of the Capacity for Cumulative Culture in Non-human Primates.

Donna Eilidh Kean

University of Stirling

Faculty of Natural Sciences

Submitted for the degree of Doctor of Philosophy.

March 2021

This thesis is dedicated to the memory of my mother,

Dr Joy Kean.

Thesis Abstract

Cumulative culture has been argued to be unique to humans, however, recent evidence from non-human animals suggests that relatively simple cases may be possible in other species. In this thesis, I explored whether non-human primates could strategically use information in a manner that may theoretically support cumulative cultural evolution. As the emergence of cumulative culture may depend on task demands, we hypothesised that favourable learning conditions may facilitate this effect. We presented monkeys with simple stimulus selection tasks, shown to produce evidence of cumulative cultural learning in human children. Success on these tasks depended on repeating rewarded and avoiding unrewarded stimuli, whose value had been revealed in a demonstration. Subjects were initially given task training to ensure that they had grasped the task contingencies and were required to meet a performance criterion before progressing. Subsequent test data was then used to simulate outcomes under linear transmission, using individual level task responses. This was done by aggregating the performance of individuals following multiple exposures to demonstrations of varying value. We tested whether, on average, exposure to higher quality information was associated with greater task success, a hallmark of cumulative culture, and whether subjects could improve upon the demonstrations they were presented with, representing increases in functionality. We also used the data to classify each subject's potential for cumulative culture according to a continuum. Several individuals displayed the capacity for cumulative culture using this method, demonstrating that it is not precluded in non-human animals. However, this may depend on significant experience of relevant cues indicating the presence or absence of rewards (as provided for our subjects in the task training). Furthermore, increases in performance were limited, apparently due to relatively low precision use of the demonstrated information. Overall, this thesis suggests that, in non-human primates, cumulative culture might be limited to minor improvements accumulated over just a few episodes of transmission, and that it is likely to be rare due to its probable dependence on favourable learning conditions. This may help to explain the apparent rarity of this phenomenon in the natural behaviour of non-human species.

Acknowledgements

My first thank you goes to my supervisor, Prof Christine Caldwell, for creating the space for such an interesting project in the first place, for your guidance, and for teaching me the confidence to think for myself. I would also like to thank my other supervisors, Dr Elizabeth Renner, and Dr Eoin O'Sullivan, and to my colleague Dr Mark Atkinson, for showing me how to do good science!

Most importantly, I am grateful to the primates who indulged my scientific curiosity and who were some of my best teachers. I would like to express my appreciation to the teams at Living Links, Blair Drummond and Kyoto University, and the organisers of the Stirling Summer Science Festival, all of whom made my data collection possible. I am especially grateful to keepers Amanda Addison, Hazel Wallace, Craig Holmes and Samir Clark for their guidance, and for sharing their knowledge. I am also indebted to Nicolas Claidière and Clare Campbell for their invaluable help with data collection.

A special thanks goes to Kate Grounds for her support and understanding, Dr Jim Anderson and Alasdair Gillies for their encouragement, and to Maxime Boürgeois and Dr Toyomi Matsuno for their patience and kindness. I would also like to give a shout out to my fantastic research assistants Chloë Wright, Amy Derrick and Guillermo Gadea.

I feel incredibly lucky for the constant support and hilarity from my PhD sisterhood: Dr Kirsten Blakey, Dr Gemma Mackintosh, Dr Juliet Roberstone and Dr Charlotte Wilks (and to our late arrival Sarah Kraemer!). For the chats, kindness, travels, advice, writing retreats, and Zoom tea breaks. I am so lucky to have gone through this experience with such strong and intelligent ladies.

A huge thank you goes to my husband, Duncan Pottinger, who has been an anchor through the most challenging times, and who also made many sacrifices for this work. This achievement is yours too. I also want to thank my wonderful friends, for their humour, kindness, and unconditional support. Particularly to Karrie Doherty, Nicola Muir, Sarah McMaster, and Scott McCartney, I am so grateful.

To my family, some no longer with us, who have been so understanding and supportive during my long academic journey: my sister, dad, step-mum, gran, nana, papa, and to my mother-in-law, Eleanor Pottinger; thank you. An extra special thanks to my aunt, Eleanor Lewis, for providing much needed sponsorship at key points throughout my academic career, and to my oldest nephew Daniel, whose enthusiasm for life and animals keeps me motivated and hopeful.

Finally, I would like to thank my mum, Dr Joy Kean, who saw the start of this journey but not the end. I could not have asked for a more inspirational role model. To you, I owe my confidence, resilience, and scientific curiosity. You were my centre and are missed every day.

Contents

Chapter 1 A Review of the Capacity for Cumulative Culture in Non-human Animals.

Introduction

Cumulative cultural evolution describes the development of increasingly complex cultural traits that is made possible by the ability to capitalise on the progress of others (Boyd & Richerson, 1996; Whiten & van Schaik, 2007). This review will highlight the evidential basis of the capacity for cumulative culture in non-human animals (henceforth 'animals'), and potential cognitive explanations for the disparity in relation to humans.

Culture, defined as socially learned patterns of behaviour that are group specific (Laland & Hoppitt, 2003), is apparent in many species. Although there may be some dispute over the breadth of evidence for culture in animals (Laland & Galef, 2009), it now appears widely accepted that many species other than humans do have this capacity (Allen, 2019). Examples have been identified in many taxa including monkeys (Panger et al., 2002; Santorelli et al., 2011); apes (van Schaik et al., 2003; Whiten et al., 1999); fish (Helfman & Schultz, 1984); birds (Aplin, 2019); cetaceans (Rendell & Whitehead, 2001); and there is even experimental evidence from fruitflies (Danchin et al., 2018) and bumblebees (Alem et al., 2016). Many cultural traditions have also been empirically demonstrated to be socially learned, to display striking stability over time, and to have no apparent ecological or geographical cause (Brown & Laland, 2003).

Cumulative culture is a sub-category of culture involving the accumulation of improvements to knowledge and skills due to a combination of social transmission and innovation, and is clear in the extensive cultural repertoire of humans (Boyd & Richerson, 1996; Tomasello, 1999a). This process has also been dubbed the 'ratchet effect' due to the directional and persistent/stable nature of these modifications (Tomasello, 1990); the terms cumulative culture and 'ratchet' will be used interchangeably throughout this thesis. By this process, cultural inheritance provides a rich foundation of knowledge that has been shaped by previous learners; this enables a short-cut to learning behaviours that have already been tried and tested, freeing up time and resources for further developments. Thus, cumulative cultural evolution allows the development of behaviours and behavioural artefacts that are beyond the inventive potential of a single individual (Boyd & Richerson, 1996). This phenomenon potentially accounts for the distinct success of humans in comparison to other animals, which can most clearly be seen in our ability to inhabit and thrive under an incredibly diverse range of environmental conditions.

Although there is evidence of cultural evolution in animals, where cultural behaviour evolves but does not increase in functionality (e.g., Lynch & Baker, 1993), there is little agreement regarding both their ability for cumulative culture, and which capacities they may lack that cause this disparity. For example, Tennie et al. (2009) argued that animals cannot accumulate culture because only behaviours they have a predisposition for, within their 'zone of latent solutions', can be socially learned, and these can also be individually invented. Complex cognitive abilities unique to the hominin lineage have been invoked to explain the variation between humans and animals, such as imitation, teaching and pro-sociality (Dean et al., 2012; van der Post et al., 2017). However, other researchers argue there is evidence for basic forms of ratcheting in animals in certain contexts, without invoking complex cognition (Claidière et al., 2014; Sasaki & Biro, 2017).

This chapter will delve more deeply into the evidence from animals, and the potential cognitive explanations for the, at least, quantitative discrepancy in cultural accumulation between humans and animals, and thus the origins of uniquely human culture. The importance of context will be stressed, as well as conceptualisation of ratcheting as a broad spectrum of phenomena with varying degrees of dependence on a wide range of different cognitive capacities (Mesoudi & Thornton, 2018).

Firstly, the cumulative culture paradox will be detailed, followed by a review of the existing evidence of ratcheting in wild non-human populations. Cognitive capacities that may explain the variation between humans and animals will then be considered, with a focus on which abilities are necessary and sufficient. The methodologies used to study cumulative culture will also be outlined, as well as their contribution to our understanding of the paradox. Finally, a proposed direction of future research will be described that addresses the limitations of previous work in identifying non-human ability for ratcheting.

The Cumulative Culture Paradox

Humans appear unique in the extent to which our culture evolves and accumulates over generations. For example, the maximum weight that can be carried by aircraft has increased over many years due to advances in science and technology (Figure 1.1), allowing more efficient transportation of goods and passengers (Bejan et al., 2014). The roughly linear trend toward increased functionality, i.e., increased efficiency in this context, displays the ratchet-like progress of many human cultural processes. In this example, later generations of airplane manufacturers likely benefited from the continuing development of, e.g., physics and engineering, relative to their predecessors. Hunter-gatherer populations also display examples of cumulative cultural evolution in, e.g., hunting tools and cooking processes (Hill et al., 2009, 2011), adding support to the seeming ubiquity of this trait across human populations (Caldwell, Atkinson, et al., 2016).

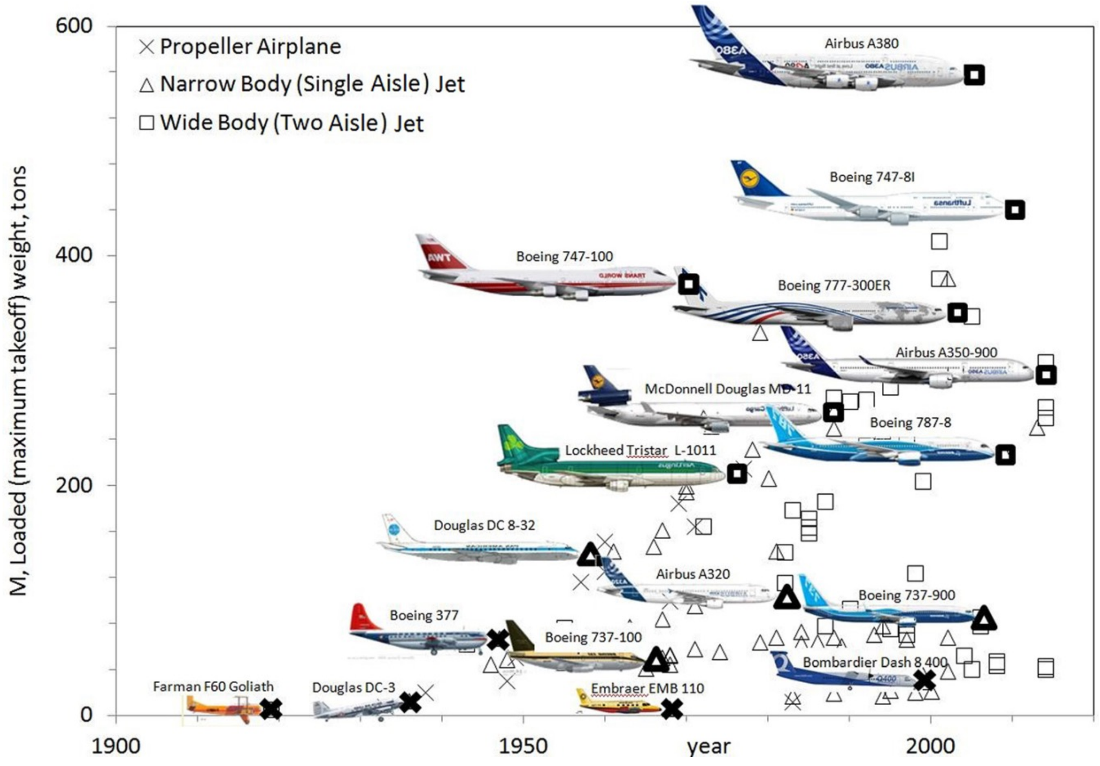


Figure 1.1. The evolution of the load capacity of airplanes. *Reproduced from Bejan et al. (2014).*

In stark contrast, and despite substantial evidence of culture in animals as noted in the introduction, evidence of ratcheting that unambiguously meets the definition of cumulative culture has not been documented. Mesoudi and Thornton (2018) identified four core conditions required for cumulative culture i) introduction of a novel behaviour, or modification to an existing behaviour, ii) social transmission of this behaviour, iii) the behaviour must deliver a benefit, an improvement in relation to existing behavioural variants, and iv) iteration of this process generating a succession of improvements. When all four core criteria have been met, this is widely accepted to determine that cultural behaviours can be classified as cumulative, therefore this definition will be adopted in the current body of work.

Some researchers impose additional requirements (see Mesoudi and Thornton's, 2018, extended criteria) to constitute the label of cumulative culture. Although these extended criteria may characterise some or even most of current human culture, these may be better conceptualised as being typical of traits that are the result of cumulative cultural evolution over many generations of transmission, rather than defining characteristics of all cumulative cultural traits (Buskell, 2020). This will be returned to in the section '*What Makes Human Culture so Unique?*'.

Some key claims for cumulative culture from the natural behaviour of animals will now be outlined. These will be considered with regard to the four criteria described above. However, no example appears to unequivocally demonstrate the full body of evidence required according to our definition, most commonly due to lack of evidence for social transmission (criterion ii), or iteration (iv).

Cumulative culture in Natural Conditions

Chimpanzees, being one of our closest living relatives, and having extensive cultural repertoires (Whiten et al., 1999), might be expected to be most likely to ratchet cultural behaviours or artefacts up in functionality. The use of tool sets to forage for honey, termites and army ants by common chimpanzees (*Pan troglodytes*) (Boesch et al., 2009, 2020; Sanz et al., 2010; Sanz & Morgan, 2009) may represent the evolution of an increasingly functional composite technology from an earlier, simpler variant. During these extractive foraging behaviours, a sturdy tool is often used to perforate the nest, then a thick stick is used to create a tunnel, and a softer stick is inserted into the tunnel to collect the food items, with the end sometimes intentionally frayed to increase the surface area. It has been hypothesised that the use of a single tool for this purpose (the only variant documented in some other chimpanzee populations, Whiten et al., 1999) was modified and improved upon as a result of social learning, and potentially separate instances of innovation.

While intuitively compelling as an argument, without historical evidence it is difficult to substantiate claims that this example of composite tool use culturally evolved from the use of single tools, or through social transmission at all. It has been argued that this behaviour is simple enough to be invented by individual learning alone. This is supported by its spontaneous, independent invention by captive chimpanzees in experiments using logically comparable tasks (Bernstein-Kurtycz et al., 2020). Similar studies have shown that other seemingly complex cultural behaviours observed in the wild can also be invented by single individuals. For example, Huffman and Hirata (2004) documented leaf swallowing, a self-medicative behaviour, being invented by individual naïve captive chimpanzees, and Tennie, Hedwig, Call, and Tomasello (2008) showed that the processing of stinging nettles in wild mountain gorillas (*Gorilla beringei beringei*) could also be individually invented by captive lowland gorillas (*Gorilla gorilla gorilla*). Such studies challenge claims that these behaviours necessarily evolved through social transmission.

Another candidate comes from a review of foraging tools cut by wild New Caledonian crows (*Corvus moneduloides*) from pandanus leaves (Hunt & Gray, 2003). These common tools are cut with the bill, and the outlines left on the leaves remain intact for around 4 years; these cut leaves were collected and analysed. Thousands of samples revealed three tool models of distinct shape, manufacturing technique and geographical distribution. The authors suggest a single origin for these tool varieties, as convergent evolution of similar tool types between groups is improbable, the distribution of each tool type overlaps, and within each type, the samples show high fidelity and an absence of intermediate designs. However, social transmission and any variation in foraging success between tool types is only inferred and has not been empirically tested, again leaving the possibility of cumulative culture an open question in this case. Moreover, at least some tool use behaviour appears to be producible by individuals raised in isolation suggesting an inherited, instinctive component to this behaviour (Kenward et al., 2005).

Schofield et al. (2018) provided longitudinal data from Japanese macaques (*Macaca fuscata*) on the spread of preparation techniques of provisioned food items (e.g., washing) and suggest they may be the result of cumulative build-up. The dispersion of various techniques for processing sweet potatoes and wheat grains were tracked through the group, suggestive of some social influence in learning the behaviours. It has also been argued that the food washing reduces parasitic load, representing a clear benefit of these developments (Sarabian & MacIntosh, 2015). However, despite the longevity of observations at this field site there is no confirmatory evidence that certain modifications developed from earlier variants.

Perhaps the best evidence comes from free-ranging bighorn sheep (*Ovis canadensis*) and moose (*Alces alces*) under semi-natural conditions (Jesmer et al., 2018). Numerous populations of these species were translocated into novel, unfamiliar landscapes, and at the time of data collection the duration of time since translocation varied across the populations. Historical populations (extant in the area for more than 200 years) were known to migrate in order to track the emergence of high-quality forage, a process known as 'green-wave surfing' which is linked to increased fitness. Levels of migration were minimal for the newly installed groups, but gradually increased to levels more closely resembling historical groups as time since translocation increased. This effect was probably driven by accumulation of green-wave knowledge which presumably increased with experience. This inferred linear relationship, as well as the length of time over which migratory rates appeared to increase, suggests that cultural transmission is the likely underlying mechanism. Although this behaviour may meet all of Mesoudi and Thornton's (2018) criteria, social transmission and knowledge increase can only be inferred from the available evidence (cross-sectional, rather than longitudinal) and is therefore circumstantial.

A major challenge for identifying naturally occurring instances of cumulative culture is the unobservable: the evolution of many behaviours leaves no archaeological trace. The low likelihood of confidently identifying social transmission under natural conditions, critical to establishing evidence of ratcheting, further compounds this issue. From a theoretical perspective, decisions regarding whether cumulative culture has occurred based on the functional complexity of a behaviour or artefact are inherently problematic, as objectively gauging what is outwith the capability of a single individual of a given species is difficult. This is especially the case at the simpler end of the spectrum of cumulative cultural outputs (where we might expect non-human examples to fall, discussed further below). Overall, even putative cases of ratcheting from observations of naturally occurring behaviours are few and far between, and do not unambiguously suggest a capacity for ratcheting in animals (Caldwell et al., 2018).

What Makes Human Culture So Unique?

A variety of traits have been proposed to explain the cumulative culture paradox. These are often suggested to be unique to the human lineage or applied by humans in a unique way. However, as already noted, cumulative culture is not a unitary phenomenon, so it is likely that the capacities required for ratcheting vary by context and task demands (Mesoudi & Thornton, 2018). This section will outline some key cognitive factors and their possible role in constraining cumulative culture in animals. It will cover imitation, teaching, innovation, social learning strategies, and working memory.

Although human culture has clearly progressed beyond such states, a focus on the essential, minimal requirements for the simplest forms of ratcheting may be most beneficial for understanding the scarcity of evidence in animals, and will therefore be adopted here. These basic forms of cumulative culture may generally be assumed to have supported the development of more complex cases (Buskell, 2020). Furthermore, aspects of the cognition required for the development of more complex traits may also rest on earlier cumulative cultural evolutionary processes (Caldwell, 2018; Heyes, 2018); however, how culture may have driven modern human cognition is beyond the scope of the current thesis and will not be a focus here.

Firstly, how social information is used has been at the forefront of explanations for the uniqueness of human cultural learning. It has been extensively argued that imitation is crucial to cumulative cultural evolution (Galef, 1992; Lewis & Laland, 2012; Tennie et al., 2009), which involves copying the form of an action through observational learning (Whiten & Ham, 1992). Imitation is a powerful social learning tool that appears integral to the stability of some cultural traits (Legare & Nielsen, 2015). Children start imitating from a very young age, suggesting it is an important developmental skill for humans (Nielsen & Mark, 2006). In contrast, it has been argued that animals do not have the capacity to use social information in this manner (Call et al., 2005).

Emulation, which occurs when the end-point of another's behaviour is copied but not the actions leading to this point (Hoppitt & Laland, 2008), has been argued to more accurately describe the upper capacity for social learning in animals (Call et al., 2005). Actions from the learner's existing behavioural repertoire are thought to be used to match the observed outcome, rather than the learning of new actions, commonly leading to significantly lower cultural transmission fidelity relative to imitation (Gergely & Csibra, 2006); this may be particularly true for more causally opaque cultural products or behaviours. This simpler form of social learning is not thought to be sufficient for cultural developments to be maintained in a population due to the lower associated copying fidelity (Tennie et al., 2009).

However, accumulating evidence suggests that several non-human species can engage in imitation, including apes (Horner & Whiten, 2005), monkeys (Voelkl & Huber, 2000), dogs (Range et al., 2007) and birds (Akins & Zentall, 1996). Although there appear to be differences in the degree to which imitation is employed between humans and other animals (Clay & Tennie, 2018; Herrmann et al., 2007; Horner & Whiten, 2005), evidence for this capacity from animals weakens claims that an inability to closely match the actions of conspecifics is a fundamental barrier (Dean et al., 2014). Moreover, empirical evidence from human adult studies suggests cumulative cultural evolution can occur without imitation (Caldwell & Millen, 2009; Zwirner & Thornton, 2015).

It is likely that task demands and difficulty are important determinants of the nature of social learning required for ratcheting in any given context (Acerbi et al., 2011). Thus, although imitation is clearly very useful, and sometimes necessary (e.g., Wasielewski, 2014), it is unlikely to be a mandatory pre-requisite for the emergence of all instances of cumulative culture. This may be particularly relevant for considering the capacities of animals, as putative examples of cumulative culture from their behaviour typically involve relatively basic traits (e.g., Boesch et al., 2009), which (as discussed above) are even argued by some to be within the species' capacity for individual learning. These are therefore less likely to require complex social learning processes. It should be noted, however, that the reverse does not necessarily follow, i.e., high levels of functionality and/or complexity should not be taken to be an unequivocal indication of the involvement of more complex social learning mechanisms.

Teaching, defined as "active information donation" (Morgan et al., 2015, p. 2) has also been proposed to play a key role in the development of cumulative culture (Dean et al., 2012). Although there is evidence of teaching from animals (e.g., Bender et al., 2009; Thornton & McAuliffe, 2006), these examples appear to represent species-typical behaviours, and constrained to specific contexts (Gergely & Csibra, 2006). This contrasts with human teaching which displays open-endedness in both context and form, and responsiveness to the needs of the learner (Caldwell et al., 2018). It has been suggested that this may be due to the capacity to represent a learner's current knowledge state (i.e., theory of mind), and distinguishes human-like teaching from animal analogues (Caldwell et al., 2018; Gergely & Csibra, 2006).

Nonetheless, evidence of teaching from animals, and evidence that it does not appear necessary for generating a cumulative effect (e.g., Caldwell & Millen, 2008) again imply that it is not a fundamental barrier. Instead teaching may become increasingly valuable or necessary as the complexity of a trait increases (Caldwell et al., 2018; Lucas et al., 2020). It has also been argued elsewhere that earlier processes of cumulative culture may have triggered the development of human specific teaching rather than vice versa (Caldwell et al., 2018; Gergely & Csibra, 2006; Lucas et al., 2020).

In combination with social learning mechanisms, innovation (broadly defined) is an integral feature of cumulative culture and has also been argued to explain the uniqueness of human culture (Legare & Nielsen, 2015). Logically, a behaviour or artefact must first arise as a novel trait before it can spread and become part of a group's cultural repertoire (Enquist et al., 2008; Legare & Nielsen, 2015). Innovation is the process by which newly learned or modified behaviour introduces a new behavioural variant into a population where it had not previously existed (Reader & Laland, 2003). There are plentiful examples of innovation from species other than humans, such as apes (Bandini & Harrison, 2020; Mendes et al., 2007) and birds (Bird & Emery, 2009; Jenkins, 1978). However, it is plausible that some of the innovations humans develop may be qualitatively different to those of animals as they are potentially more insightful and deliberate, as opposed to serendipitous discoveries based on trial and error that may predominate amongst animal innovations (Caldwell, Cornish, et al., 2016; Legare & Nielsen, 2015). It has been argued that the necessary cognitive capacities likely vary across classes/types of innovation (Legare et al., 2015; Reader et al., 2016), again highlighting the importance of context in determining the emergence or otherwise of cumulative culture.

Thus, across the suite of capacities discussed which have each been proposed to underpin human cumulative culture, a common theme emerges, suggesting that it is not the presence or absence of a single specific skill or ability (Laland & Seed, 2021), but rather the way in which information is used that may support human cumulative cultural evolution at a fundamental level.

Thus, a more general, foundational capacity to differentially determine when and how to use social information may underlie the development of even the simplest instances of cumulative culture (Heyes, 2016; Laland, 2004). 'Social learning strategies' are rules which cause learners to be selective about when, from whom, and what to socially learn (Price et al., 2017). These can involve decisions such as whether to learn socially or individually, or strategies such as copy-the-most successful individual (Laland, 2004). It has been formally demonstrated that the ability to socially learn, regardless of the mechanism, would not be adaptive without some selectivity (e.g. Rogers, 1988). Thus, social learning strategies offer rules that can increase the adaptiveness of learning (Enquist et al., 2007; Miu et al., 2020; Rendell et al., 2010).

Although social learning strategies are common in the behaviour of non-human species (e.g., Battesti et al., 2015; Grampp et al., 2019; Horner et al., 2010; Kendal et al., 2009), it has been suggested that humans have a special type at their disposal which we do not see examples of in animals (Heyes, 2016; Dunstone & Caldwell, 2018). Heyes (2016) referred to these as 'explicitly metacognitive social learning strategies', and they are proposed to involve higher level cognitive processes that allow the implementation of learning rules that are more flexible, specific, and precise, potentially optimising the efficacy of human social transmission. The information transfer resulting from such high level strategic social learning may support the emergence of human-like cumulative cultural evolution. This is due to the ability to consciously represent the rules, and therefore make reasoned decisions about how to use information based on an understanding of one's own and/or others' minds. Importantly, this may even apply to the general use of information, regardless of whether it has been derived from a social source or individual exploration (Dunstone & Caldwell, 2018). The strategies of animals (and those used by humans much of the time) are likely based on fast, non-conscious/automatic mechanisms, also called System 1 processes, that are typically learned by association. However, explicitly metacognitive social learning strategies are proposed to rely on slow, deliberate, conscious cognitive processes (System 2) and are thought to be employed in circumstances where System 1 strategies are inadequate. Such strategies might also sometimes involve an understanding of the mental states of others (i.e., theory of mind), and therefore allow reasoning about how to learn from the behaviour of others (Dunstone & Caldwell, 2018). Evidence for both theory of mind and metacognition from animals is scarce, with even the best examples (e.g., Beran et al., 2015; Krupenye et al., 2016) potentially being explained by associative learning rather than metacognitive understanding.

It is likely that not all forms of ratcheting require social learning strategies that are based on meta-cognitive or theory-of-mind reasoning. However as explicit strategies regarding information use more generally permit greater precision and flexibility (e.g., higher fidelity copying of preferential, and consistent avoidance of sub-optimal behavioural variants) these may have the potential to generate behaviours that are closer to optimal across a variety of contexts. This potentially underlies some of the distinctiveness of human culture at a fundamental level.

Overall, requirements for cumulative culture may have been overestimated by focussing on late appearing or complex traits as well as a narrow range of contexts (Buskell, 2020; Shettleworth, 2010). It is likely that at least some forms of ratcheting do not require action-copying, teaching, complex innovation, or meta-cognitive reasoning, particularly for embryonic examples of cumulative culture (i.e., less well-developed traits), even though these skills may be implicated in much of modern human culture. Moreover, we have also cited examples from the behaviour of animals consistent with these traits, demonstrating that they are, in and of themselves, not unique to humans in any case.

However, a general (i.e., not specific to the social domain) ability to explicitly represent and use learned information may be an important factor for the type of high-precision, selective information use necessary for any cultural accumulation to take place. Although cumulative culture requires social information use by definition, the extent to which simple forms rely on specifically social skills alone is unclear; this will be returned to in Chapter 2. It is plausible that general cognitive mechanisms (such as inhibition, attention, working memory) may also play a key role. Perhaps, the learning strategies employed by animals may broadly involve use of information in a manner that is simply too imprecise or unselective to generate a ratchet effect.

Working/Short-term Memory

Having emphasised in the previous section the importance of context in determining which abilities may be necessary for cumulative culture, the following section will consider the case of working/short-term memory as an example to illustrate this.

In humans, products of cumulative cultural evolution are apparent in an extremely wide variety of contexts and domains. For example, humans have the ability to accumulate cultural modifications in fields such as hunting, which provides direct fitness benefits, but also those that are adaptive in a broader or more indirect sense, such as mathematics and computer science. In contrast, the clearest examples of cumulative culture in animals appear restricted to the fine-tuning of pre-existing natural tendencies (Jesmer et al., 2018; Sasaki & Biro, 2017). Several capacities may explain this relative open-endedness in the cultural behaviour of humans that appears unmatched in animals. For example, executive functions may play a key role, as capacities such as attentional control and inhibition may broadly affect the use of information (Coolidge & Wynn, 2001).

It has been suggested that working or short-term memory may have been fundamental in the development of cumulative culture in humans (Balter, 2010; Read, 2008; Wynn & Coolidge, 2007, 2011). Short-term memory involves the capacity for storage of a small amount of information in a readily available state, and working memory involves such storage, as well as the capacity to manipulate the information (Cowan, 2008; Diamond, 2013).

Contexts which place demands on memory may vary the potential for ratcheting, depending on the ability of the population. For example, if the activity of an individual leaves behind physical cues (that have some longevity) this may aid subsequent use or manipulation of this information by social learners through relaxing memory pressures. This may hold even in the absence of witnessing the behaviour itself. For example, puncture holes in milk bottle caps left by previous foragers may provide vicarious information about the value of the bottle's contents (Fisher & Hinde, 1949; Sherry & Galef, 1984), and ants leave pheromone trails that provide navigational and regulatory information that are used by conspecifics to guide foraging behaviour (Czaczkes et al., 2015). Tennie et al. (2009) also argued that such cues are instrumental in the development of cultural behaviours, e.g., nutcracking, in chimpanzees (but only within their zone of latent solutions) due to the information they carry regarding environmental affordance.

In contrast, behaviours that leave little trace may not only require active attention to be paid in real time for learning to occur, but also storage of the relevant information for subsequent use, for example, in many social conventions of humans. Socially learning causally opaque tasks may also require observation and storage of the necessary actions even if some remnants of the behaviour are left behind, as observation of the traces alone may not elucidate their purpose.

Typical social learning experiments, such as artificial fruit paradigms (e.g., Marshall-Pescini & Whiten, 2008) provide demonstrations with only transient cues to task solutions, and the goal is typically to repeat this demonstrated behaviour. As the demonstration leaves no lasting trace, the behaviour must be stored in memory, creating cognitive demands. In contrast, contexts in which the best candidates for ratcheting have been found in nonhumans (Jesmer et al., 2018, discussed above; Sasaki & Biro, 2017, described below) involve minimal memory load, as acquiring improved variants is likely an incidental outcome of other behavioural predispositions, i.e., herding and flocking.

As such, greater working/short-term memory capacity might be a non-social moderating factor in the emergence of cumulative culture, whereby greater pressure on this capacity may increasingly restrict the contexts in which a ratchet effect is possible. Empirical support for this theory has been documented by Wilks et al. (2021). They investigated this possibility in children and found evidence that conditions of greater memory load inhibited emergence of a cumulative effect in younger age groups. Fay et al. (2019) also suggested that increased working memory pressure may have inhibited the development of cumulative culture in a transmission chain experiment with human adults, compared to conditions involving (inferred) lower taxation of working memory. It is plausible that this may also hold true for animals.

Overall, it is unlikely that a single cognitive capacity is required for cumulative cultural evolution, but that the required capacities vary by context. Specifically social abilities may be prime candidates given that social transmission is a requirement for ratcheting; however, we have argued that relatively well-developed executive functions (e.g., working/short-term memory) may also be required for the learning or innovation involved in simple instances of cumulative culture. At the level of minimum cognitive requirements, inexact or inconsistent information use potentially constrains accumulation in non-human cultures.

How has Capacity for Cumulative Culture been Operationalised?

Various methodologies have been employed to assess capacities for cumulative culture. This section will outline three key approaches, namely adaptive solution-switching studies, closed groups methods, and generational replacement designs. The relative utility for testing for cumulative culture in animals will be described for each of these methods.

Adaptive Solution-Switching Methods

Early studies of cumulative culture assessed the capacity to abandon established techniques, typically for acquiring food rewards, to adopt more efficient methods. This type of flexibility is thought to be integral to the process of cumulative culture to allow the spread of novel behavioural variants or modifications that confer some advantage relative to established variants (Davis et al., 2016). As such, the capacity to recognise and adopt beneficial techniques in response to social information was hypothesised to underlie ratcheting (Marshall-Pescini & Whiten, 2008).

This was typically assessed using artificial foraging tasks, and usually with great ape subjects; see Whiten et al. (2009) for a review of such studies in chimpanzees. An early study employing this paradigm involved initially demonstrating a simple dipping method for accessing honey from a foraging device to chimpanzees (Marshall-Pescini & Whiten, 2008). This required sliding a small trap door open on the lid of an artificial foraging device and using a rod as a dipping tool to acquire relatively small amounts of honey through this opening. This was readily acquired by five of the subjects. A more efficient and intricate technique was then demonstrated to these five chimpanzees on the same apparatus, which built upon the simple method: the same rod tool was firstly used to probe a bolt which unlocked the lid of the device; the small trap door was then opened, and the rod was inserted (as in the simple method). However, rather than dipping the rod in the honey, it was used as a lever to prise the whole lid open, providing access to the entire deposit of honey. Thus, the more efficient method built upon the simpler method, involving greater action sequence complexity, but a more efficient means of gaining reward. The researchers then observed whether the former method was abandoned, and the more effective variant adopted. Only one individual displayed this solution-switching behaviour, while the other four continued to employ the more basic dipping method. This conservatism was suggested to represent a barrier to the uptake and spread of beneficial modifications in animals.

However, studies using similar paradigms have found mixed results. Some have found evidence of relinquishing sub-optimal behaviours for more efficient variants (e.g., Van Leeuwen et al., 2013; Yamamoto et al., 2013), whilst other results show more conservatism (e.g., Bonnie et al., 2012; Hrubesch et al., 2009). This variation may be understood by the wide variety of tasks employed

(Davis et al., 2016), as some contexts may have been more conducive to learning for the study species than others.

However, the outcomes of single transmission events, or single instances of social learning, do not in any case directly evaluate the capacity for ratcheting. The build-up of beneficial modifications to solutions over multiple learner generations is required to empirically demonstrate a ratchet effect (Caldwell & Millen, 2008). Thus, even in cases where subjects did successfully switch to a more efficient technique, this does not provide positive evidence that such behaviour would support ratcheting over repeated transmission (Caldwell et al., 2020). Furthermore, negative evidence resulting from this paradigm does not sufficiently demonstrate that cumulative culture is precluded but may simply represent context-specific difficulties.

Closed Groups Methods

Relative to solution switching designs, closed group studies have a greater capacity to test the potential for beneficial modifications to accumulate within a population. This methodology involves a fixed group of participants (i.e., group membership is constant, with no replacement) that are typically repeatedly exposed to a task in an open testing environment (Mesoudi & Whiten, 2008). Whether solutions to the task can be ratcheted up, and whether this is aided by the use of social information, can then be assessed.

Using this method, Dean et al. (2012) aimed to evaluate whether imitation and teaching (amongst other factors) were conducive to generating increasingly efficient task solutions. Groups of children, chimpanzees and capuchin monkeys (*Sapajus apella*) were presented with an identical puzzle box task involving three levels of increasing difficulty. In one condition, levels two and three could not be attempted until the subsequent level(s) had been completed. Level one required a single action (slide door) to access a low-grade reward; the second level added another action (push button and slide door further) to receive a higher value reward; and level three required a different action (twist dial and slide door even further) to access the highest value reward.

The children were the highest performing, with at least two members out of five of the groups (from eight groups in total) found to solve the puzzle box to the most difficult level, compared to only one individual chimpanzee and no capuchins. Teaching and imitation were linked to the successful behaviour of the children and argued to contribute to their relative success, as there was little suggestion of social learning in the less productive non-human groups.

However, in such closed group designs, it is not possible to determine whether increases in success are necessarily due to social transmission, or increased experience with the task as a result of individual learning (Caldwell et al., 2020). The latter could also result in improved performance over repeated task exposure due to carry over effects. As the iteration of social information transfer is core to cumulative culture (Mesoudi & Thornton, 2018) this is an important distinction and has implications for the suitability of these methods for identifying ratcheting. Indeed, individually tested children of the same age were found to solve the task to all three levels in a recent follow up study (Reindl et al., 2020), suggesting that individual learning alone can generate this level of success. Reindl et al. (2020) suggest that social facilitation may have encouraged engagement with the task in Dean et al. (2012) rather than social learning of the task contingencies. The design of these studies does not make it possible to determine whether access to information from another learner's attempts would benefit a fully naïve participant, due to the fixed structure of the group.

Furthermore, an inability to solve the task used by Dean et al. (2012) to the highest level does not necessarily mean that cumulative culture is impossible for these species' in other contexts involving different task demands. Consistent negative results across a wide breadth of opportunities for ratcheting would be necessary to robustly substantiate this claim (Caldwell et al., 2020).

Generational Replacement Methods

Generational replacement methods systematically remove and replace participants in a sequence, directly addressing the first of the limitations mentioned above in relation to closed group studies, i.e., concerning the benefits of social information to newcomers. Generational replacement methods involve chains of learners, capturing the inter-generational turnover that is fundamental to cumulative culture, as well as monitoring any increases in performance. As such, these designs can more convincingly demonstrate the capacity for ratcheting in experimental conditions through dissociating individual exposure times from the extent of prior transmission.

Generational replacement designs involve transmission chains of *n* participants, where the replacement of participants creates a new learner generation. The first generation attempts a task from a position of naivety. Generation 2 is then exposed to some information produced by Generation 1 (e.g., process information or a product resulting from their attempt) and participates in the same task. Generation 3 is then exposed to information from Generation 2's attempt, and so on for *n* generations. This process where the previous generation's output becomes the input for the following generation is the crux of this design: each chain (from Generation 1 to G_n) is a simplified model of the inter-generational learning process in the real world. For cumulative culture to be inferred from a particular chain's performance, the success (and hence also information received) of later generations must be of higher value compared to earlier generations, according to an objective measure of success.

These designs can involve unidirectional information only, with no overlap between the learning of each generation, i.e., transmission chains, or some bidirectional transfer of information where the learning of generations can be influenced by multiple previous generations, i.e., replacement designs (Caldwell et al., 2020; Mesoudi & Whiten, 2008).

Using a replacement design, (Caldwell & Millen, 2008) confirmed the utility of these methods for eliciting a ratchet effect in human adult populations. They asked participants in some chains to create paper aeroplanes that could be thrown as far as possible, and others to build the tallest tower they could from uncooked spaghetti and clay. Across twenty chains of ten generations each, success increased as the chains progressed, with final generations being more successful than the earliest (Figure 1.2). These tasks allowed success to be objectively quantified (distance flown and tower height). Each generation also had a comparable level of exposure to the task and to information about others' performances, thus constituting a rigorous evaluation of cumulative progression.

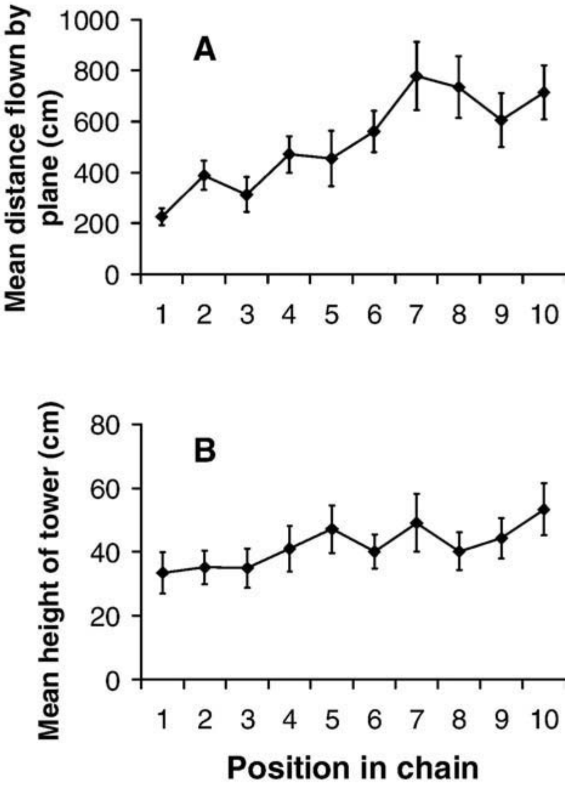


Figure 1.2. Success over generations in transmission chains where the aim was to build paper planes (top) and build spaghetti towers (bottom). *Reproduced from Caldwell and Millen (2008).*

Controlled application of these methods can also tease out which factors are necessary for cumulative culture by manipulating the conditions of the tasks. As mentioned in the previous section, the necessity of imitation and teaching for ratcheting was thrown into question by a generational replacement experiment with adults (Caldwell & Millen, 2009). Participants were given the paper aeroplane task described above and assigned to one of three conditions, each of which provided different types of social information about the aeroplanes of previous generations: i) the building process (allowing imitation), ii) the finished artefact (allowing emulation) and iii) the opportunity for previous participants to help later generations (allowing teaching). It was found that emulation alone was sufficient to produce a cumulative effect. This conclusion has been challenged by Tennie et al., (2012) who claimed it is likely participants would have previous experience with the task, and that the skill was probably initially learnt by imitation. However, these results were replicated using a rice carrying task that participants were unlikely to have experience of (Zwirner & Thornton, 2015).

Generational replacement methodologies have also been used with animal species to assess whether cultural transmission may be possible in experimental settings. For example, a transmission chain experiment with Guinea baboons (*Papio papio*) aimed to replicate methods used in human language evolution experiments (Claidière et al., 2014). In a key human language evolution study, Kirby et al. (2008) studied the cultural evolution of systematicity in a communication system, employing an artificial language task with adult humans using the generational replacement design. The first generation were asked to learn randomly generated combinations of nonsense labels that referred to shapes. Test trials required the participants to reproduce these labels from memory when subsequently presented with the shapes they had seen, as well as novel shapes that they have not encountered before. Subsequent learners were tasked with replicating the output of the previous generation, i.e., the labels produced during the test trials. Over several generations of transmission, the labels became systematically structured and therefore more learnable/easier to copy.

Claidière et al. (2014) also showed the development of systematic structure from a simplified, abstract adaptation of these studies. The baboons were shown a grid of sixteen square stimuli on a touchscreen computer, four of which were red and twelve of which were white for 400ms (Figure 1.3). All stimuli then became white, and the subjects were required to select at least three out of four of the previously red squares to attain a food reward. Transmission chains were created by using the output of Generation 1 (the squares they touched) as the input for Generation 2, and so on.

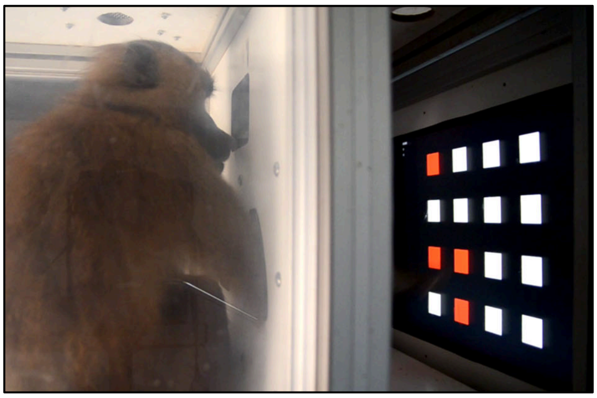


Figure 1.3. Baboon taking part in the touchscreen stimulus selection task. *Reproduced from Claidière et al. (2014).*

An increase in transmission fidelity was found, presumably attributable to the fact that the task was made easier for later generations by the patterns becoming increasingly clumped over transmission as a result of accumulated copying error.

This example illustrates how a generational replacement design can be used to illustrate cultural evolution in a broader sense. However, cumulative culture as we have defined it (and consistent with others' definitions, e.g., Mesoudi & Thornton, 2018) equips later generations with better quality information conferring an advantage relative to earlier generations performing the same task. Such an effect was not demonstrated in the study just described, although the work clearly bears strong similarities to much of the work carried out in the current thesis (and will be returned to in Chapters 4 and 5).

In another key study, the efficiency of the routes of homing pigeons (*Columba livia*) using replacement methods were monitored using GPS tracking (Sasaki & Biro, 2017). Firstly, Generation 1 stabilized a homing route by flying it twelve times. Individuals placed in this first generation were then paired with a naïve pigeon (Generation 2) and they also flew together twelve times. The first-generation bird was then removed, and Generation 2 was paired with a naïve bird (Generation 3) who were then released for twelve flights; this was repeated for five generations in total. Unlike control conditions (ten solo birds, and ten fixed pairs of birds also released 60 times each), only the generational replacement chains showed a marked increase in navigational efficiency over the course of the chains of learners. The authors suggest that new generations altered the routes, pruning them of inefficiencies, and claim that this represents the capacity for ratcheting as a result of collective intelligence. It seems that in this case, the cumulative effect is therefore an emergent property of collective decision making based on the natural propensity to flock. Thus, although the effect found appears to meet Mesoudi and Thornton's (2018) criteria, this ability is unlikely to be generalizable to other circumstances. For example, it is unlikely that a cumulative effect would be found in a context in which the matching of behaviour would not be expected anyway, for reasons other than information acquisition (in this case occurring as a consequence of tendencies for flocking).

Overall, due to opportunities to control the learning conditions and observe information transfer, generational replacement methods are currently the most effective method by which capacity for ratcheting can be assessed. However, studies with animals to date are limited in their ability to unambiguously demonstrate a cumulative culture effect due to issues in task design. The following section will outline an alternative method for inferring ratcheting potential.

The PFR Method

As cumulative culture likely involves a broad range of processes and dependencies, rather than describing any particular population as 'having' cumulative culture or not, ratcheting capacity probably exists along a continuum of ability. Its emergence is plausibly influenced by an interaction between context and cognitive architecture, as discussed above, therefore asking under which circumstances it might be observed may be a more prudent approach.

As there is currently no unambiguous demonstration of cumulative cultural evolution in non-human species, suggesting it may be a fragile phenomenon, attempts to identify this process may benefit from creating optimal conditions to facilitate its emergence. Whilst also providing the best opportunity for identification, adopting this approach across a variety of contexts may also provide stronger negative evidence in the event of null results (Caldwell et al., 2020); claims of a lack of cumulative culture in non-human species often involve learning in a relatively narrow context, and it is unclear whether this results from a lack of capacity for ratcheting, or task-specific difficulties (e.g., Vale et al., 2020). Minimisation of any cognitive burdens, such as memory load, as well as simplistic task designs (whilst maintaining the necessary features to test for cumulative build-up) may be key to creating these conditions.

As outlined above, generational replacement methods do have the power to effectively demonstrate capacity for cumulative culture (given positive results). However, due to significant logistical and organisational demands, these methods are particularly challenging to implement with certain populations, e.g., animals and young children. For example, Caldwell and Millen (2008) recruited 100 participants for each demonstration of cumulative culture involving the construction of paper planes and spaghetti towers, discussed above. To manipulate the opportunities for social learning (emulation, imitation and teaching, discussed above), Caldwell and Millen (2009) recruited 700 participants, involving 100 individuals in each of the seven learning contexts they created. The already large sample sizes required are thus multiplied when attempting to compare across varying learning contexts. The ability to investigate the range of conditions under which cumulative culture may be observed (or not) may be of particular consequence for non-human populations where the range of contexts may be especially constrained. However, clearly, these numbers would not be viable for many non-human species, and particularly for those, such as great apes, that are classified as endangered.

Horner et al. (2006) also displayed the organisational difficulties in constructing transmission chains with chimpanzees in a study of copying fidelity. They created only two chains involving four and five generations, each of which were seeded with one of two techniques for retrieving a food reward from a foraging device; whether the seeded method was faithfully transmitted along each chain was then observed. Part of this process involved creating conditions to allow each new generation to closely observe the performance of the previous generation, allowing for social learning to occur. This required considerations such as dominance relationships between the demonstrator-observer pairs at each point of generational turnover, so that, e.g., the observer would not prematurely displace the demonstrator. Their design also required the demonstrator to be coaxed out of the research room with a food reward when the demonstration phase was complete, in order to allow the observer the opportunity to interact with the apparatus. Despite best efforts, aggressive behaviour and noncompliance interrupted both chains, requiring some inter-generational transmission episodes to be re-attempted. This study displays the substantially greater challenges in using these designs to test non-human primates relative to humans.

To address these issues, the PFR (i.e., potential-for-ratcheting) Method retains the key features of generational replacement designs whilst minimising the methodological burdens (Caldwell et al., 2020). A full description can be found in Caldwell et al. (2020), however the key points will be described here.

The crux of this method is the use of relatively large quantities of data from individual subjects to infer the capacity for cumulative culture at the individual level. To do this, subjects are repeatedly exposed to task demonstrations which involve information of varying value. This exposure to the range of scores (that could be attained on the task) makes it possible to simulate the improvement in task performance across generations of learners in typical generational replacement experiments. Thus, this feature allows assessment of any positive relationship between increasingly useful information and task success, a key indicator of cultural accumulation.

The potential benefit of information transfer can subsequently be inferred through assessing how the respondent’s performance compares to that demonstrated to them, on average. To adopt the example provided by Caldwell et al. (2020), consider a stimulus selection task involving 16 stimuli, and demonstrations where four of these stimuli are selected (Figure 1.4, lower section). This can either be performed by an actor, or vicarious information can be provided by an animation or timed display. Within the grid there are four randomly dispersed rewarded target stimuli (the green squares), and the rest are unrewarded (red squares) so that demonstrations could score between *0* (no targets found) to *4* (all targets found). Optimal use of this information would either match or improve upon the demonstrated performance by reselecting target stimuli and avoiding non-targets by choosing alternatives.

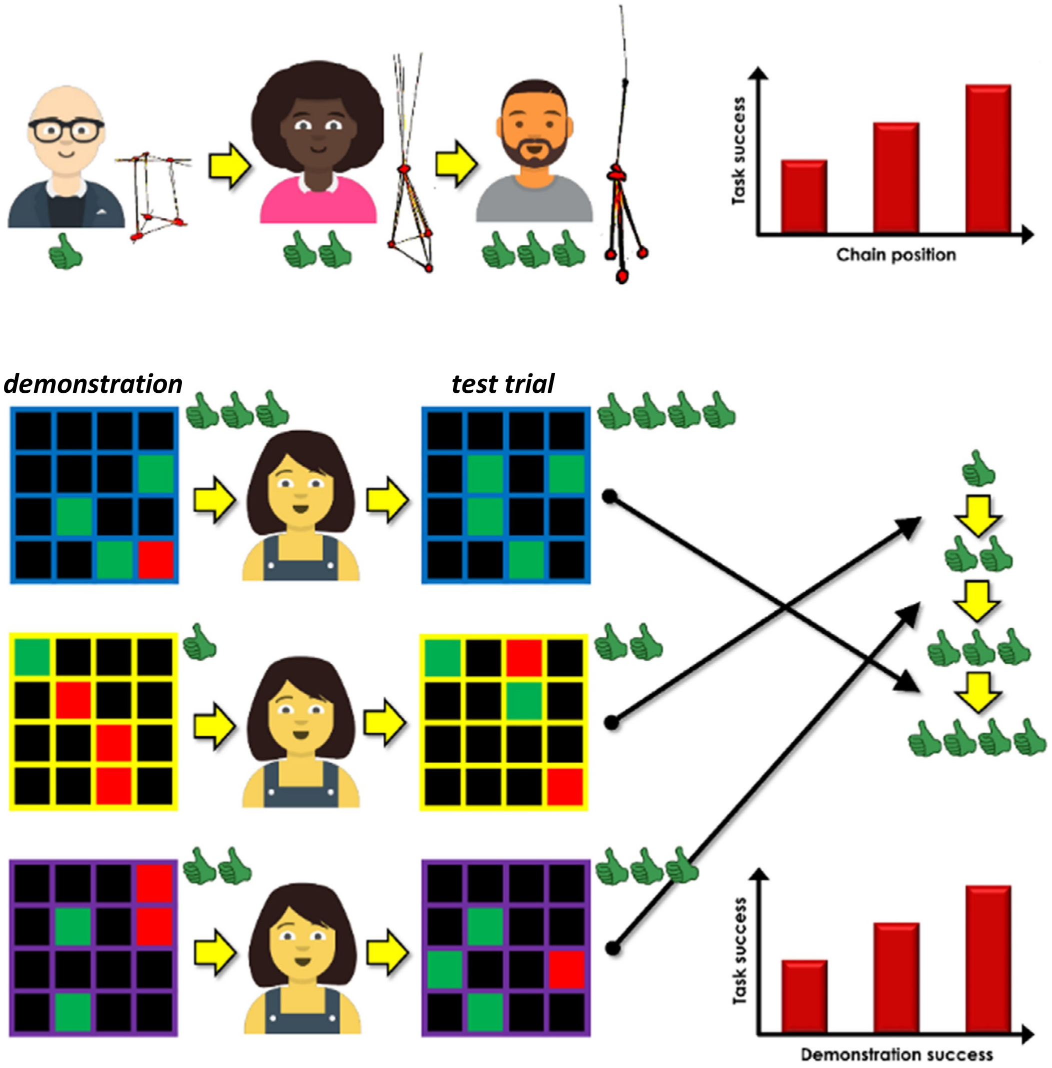


Figure 1.4. ***An illustration of a stimulus selection task that could be used with the PFR Method.*** *Arrows indicate the direction of the transmission of information, i.e., the information input and output for each participant. Adapted from Caldwell et al. (2020).*

Demonstrations displaying the location of one target (and three non-targets) reflects performance that might be expected from a naïve subject, i.e., chance level, whereas higher scoring demonstrations reflect the output of a more proficient individual who had themselves learned from a previous generation. Thus, different values of information correspond to theoretical positions in a chain, such as the spaghetti towers experiment developed by Caldwell and Millen (2008); see Figure 1.4, top section.

Multiple exposures to each potential demonstration score are possible using this task design as randomisation of the location of targets between trials would eliminate carry over effects; thus each presentation would constitute a fresh problem. Aggregating performance in response to each demonstration score can then simulate whether improvements to these scores would theoretically be expected across a chain of transmission; such improvement would be possible provided the correct strategy was employed consistently, and given enough trials. For example, consider the optimal strategy was implemented each time a demonstration scoring *2* was presented. Multiple exposures to such information would typically generate a higher score overall, as two selections would retain the targets displayed, and the remaining two would be expected to occasionally chance upon a hidden target(s).

As such, instead of independent participants comprising a chain of learners, the outcome of this process can be derived from individuals, or very small groups, which may otherwise not generate enough power due to low sample sizes or could not be arranged into a chain due to logistical barriers. However, any task used to implement the PFR Method must allow objective measurement of performance and be free of carry over effects, otherwise prior experience would obscure any differential effect from demonstrations of differing value. The example described above meets these criteria. Moreover, although the vicarious demonstrations provided in this example may not be described as ‘social’, they are vicarious in the sense that they were not generated by the learner. The mode of presentation of information could in fact be another contextual factor that may influence information use. However, previous work from our lab group has found minimal variation in performance between social demonstrations that were performed by the experimenter and those that provided vicarious information only using a similar task to that described here (Renner et al., 2021).

The feasibility of the PFR Method has the potential to elucidate the capacity for cumulative culture without the burden of coordinating multiple participants. The simplicity of the type of task described above may further create a context by which, in principle, a cumulative culture effect could be elicited in animals. It also allows manipulation of learning conditions; for example, visible or enduring cues could be provided during test trials to reduce the memory load involved in memorising demonstrated solutions. This logic forms the basis of the task used in Chapters 4-5 (and validated in Chapter 3) to assess cumulative culture capacity in non-human primates.

Thesis Goals

The current chapter has summarised the existing literature regarding the evidence for cumulative culture in animals. Overall, the capacity for ratcheting in non-human species remains unclear. However, evidence suggests that this ability may not necessarily be precluded in non-human populations, but very rare, as optimal conditions may be required. This section will outline the contributions of the empirical work in this thesis in relation to deepening our understanding of this field.

The ultimate goal of this project was to evaluate whether non-human primates have the capacity for cumulative cultural evolution using the PFR Method, described above.

Chapter 2 firstly assessed whether the type of strategic information use that may underlie cumulative cultural evolution is available to a non-human primate species. It also capitalised on the opportunity to explore a common assumption of many theories of comparative social learning, which is that information acquired socially may be treated differently from information acquired individually. We were able to directly compare the way in which information was used, depending on whether it has been acquired from social observation or individual exploration.

Chapter 3 validated the touchscreen-based task we aimed to use to assess ratcheting in animals, which followed the logic set out by Caldwell et al. (2020) and described in the previous section. Proof of principle of the PFR Method had been presented in a study which demonstrated the ratchet effect in young children (Wilks et al., 2021). Therefore we initially sought to replicate these findings with children in order to determine the validity of our task in detecting a similar effect.

Finally, Chapters 4 and 5 sought to evaluate the capacity for cumulative culture in two monkey species, using the PFR Method to infer the outcomes of cultural transmission.

Overall, these studies explored whether cumulative culture may be theoretically possible in non-human primates by assessing the capacity to use information in an optimal manner. These studies contribute to the existing literature by providing the first attempt to test whether two monkey species can strategically learn in a manner that supports cumulative improvement.

Chapter 2 Capuchin Monkeys Learn to Use Information Equally Well from Individual Exploration or Social Demonstration.

Contributions

Donna Kean, Christine Caldwell, Elizabeth Renner and Mark Atkinson contributed to study conception, design and material preparation. The code for the task was written by Mark Atkinson. Data collection was performed by Donna Kean and Elizabeth Renner. The data analysis was performed by Donna Kean with contributions from Mark Atkinson and Christine Caldwell. The chapter was written by Donna Kean and edited by Christine Caldwell.

Introduction

The ultimate goal of this thesis was to test non-human primates' ability to use vicarious information selectively in order to effectively guide their behaviour. This ability may be a fundamental prerequisite for cumulative cultural evolution (Laland, 2004).

The generational replacement design is a well-established method for evaluating cumulative cultural evolution in the lab by creating chains of learners (reviewed in Chapter 1, section '*How has capacity for cumulative culture been operationalised?').* It tests the ability of the participants, arranged into a sequence of 'generations', to improve upon the success of earlier generations after being exposed to their solutions. For example, for successive improvement to emerge from transmission chains using the spaghetti towers task (Caldwell & Millen, 2008), agents are required to select, retain and accumulate useful construction information (e.g., triangular bases are sturdy) and disregard variants leading to poor performance (e.g., liberal application of modelling clay at upper tower sections causes instability). This design models the intergenerational process of descent with modification over learner generations by which cumulative cultural evolution occurs.

In order to adapt this methodology for non-human primates, we aimed to develop a task to meet their competencies whilst maintaining the crucial design features of the studies employed with humans. Without the verbal instruction possible with human participants, the first step toward this goal was to train subjects to use the underlying task cues. Once the cues could be used with proficiency, the task could be gradually expanded to a level of complexity that sufficiently addressed capacity for the accumulation of expertise over generations, as a result of social transmission (Chapters 3-5). This is also sometimes described as the "ratchet effect" (Tomasello, 1990, p. 305). This chapter will outline this initial training process with non-human primates, with the overarching goal of scaffolding their learning to effectively assess their potential for ratcheting using similar cues at a later stage (Chapters 4 and 5).

In conjunction with the training, we also capitalised on the opportunity to test a fundamental question relevant to cumulative culture theory: potential differences between humans and other primates in their use of socially vs individually acquired information.

Human-unique social learning mechanisms have been proposed to explain the relative complexity of human culture. Specifically, humans are argued to rely more heavily on social learning and to engage in higher fidelity copying of behaviour than non-human primates (Boyd & Richerson, 1996; Montrey & Shultz, 2020; Visalberghi & Fragaszy, 1990). The ability to closely match observed behaviour is thought to have allowed the retention of beneficial information over generations and reduced backward slippage in human societies, leading to the emergence of cumulative culture (Lewis & Laland, 2012; Tennie et al., 2009). Moreover, cognitive mechanisms have been proposed to have evolved for learning and processing specifically social information in humans (e.g., Cosmides & Tooby, 2015; Herrmann et al., 2007). In contrast, non-human primates are argued to depend primarily on individual learning (Tennie et al., 2009).

However, although social learning is by definition necessary for culture to accumulate over generations, the extent to which this process is based upon learning that is inherently specific to the social domain remains unclear (Osiurak & Reynaud, 2019). Adequate comparisons between social and individual learning which could evaluate this question are lacking, despite strong interest in the role of social learning processes in distinctively human culture.

The current study replicated the methods of Renner et al. (2019) who investigated this hypothesis using a population of squirrel monkeys. Renner et al. (2019) developed a win-stay, lose-shift (WSLS) task that tested how information that was either rewarded or unrewarded was used when the information source was either a social demonstrator, or the subject. Atkinson et al. (2020) also ran a version of this task with young children.

The WSLS structure of the task required subjects to repeat selections that were rewarded (‘win-stay’) and avoid those that were unrewarded (‘lose-shift’). Problems involved an information trial where a selection was made between two (Stage A) or three (Stage B) stimuli on a touchscreen computer. This information trial was performed by the experimenter in the social condition or the subjects themselves in the individual condition, but crucially, identical information was provided across both conditions. Test trials immediately followed whereby the subjects selected from the same set of stimuli. If the information trial was rewarded the successful strategy was to repeat the selection made. If it was unrewarded, the successful strategy was to deviate from the demonstrated selection and choose an alternative. Thus, sometimes the optimal response was to copy and sometimes it was to explore.

The squirrel monkey (*Saimiri sciureus*) population tested by Renner et al. (2019) performed poorly on this task meaning hypotheses regarding information source could not be assessed. The current study aimed to address this issue by testing a population of tufted capuchin monkeys (*Sapajus apella*), where there is some reason to expect greater success, discussed further below.

Previous methodologies testing non-human primate social learning have arguably failed to effectively compare social to individual learning. These studies typically involve a social condition where a vicarious demonstration provides information about how to successfully retrieve food from a puzzle box, often through use of a tool, and success is defined by copying this solution (e.g., Davis et al., 2016; Marshall-Pescini & Whiten, 2008; Tomasello et al., 1987). This is often compared to an individual exploration condition where no such information is provided but where the subject can freely engage with the apparatus. Evidently, there is no scope for copying in the latter condition, but success requires spontaneously solving the puzzle box during naïve exploration. Thus, there are a number of significant disparities in the information provided between the social and individual learning conditions. In one such tool-use study, Nagell et al. (1993) provided a demonstration of a target behaviour to chimpanzees (*Pan troglodytes*) in a social condition, where a rake-like tool was manipulated to pull food within a reachable distance. An individual control group were given no demonstration but were simply allowed to explore the apparatus. The social group were more successful when subsequently attempting the task, compared with the individual group who were required to spontaneously discover the behaviour. These unequally matched conditions are problematic for drawing sound comparisons regarding the use of information accessed socially or individually.

A group of studies have used ghost controls in an effort to establish informational equivalence with a social condition to address this limitation. Ghost demonstrations match the object movements during the demonstration performed by a social model, but without the perceivable involvement of an actor. For example, Hopper et al. (2007) compared whether chimpanzees' ability to solve a tool-use task would vary by whether the demonstration was performed by a conspecific or not. Two separate ghost controls involved a demonstration where either the apparatus was automatically triggered, or a tool seemingly operating of its own accord (using hidden fishing wire) successfully released the food reward. They found that the social demonstration group were significantly more successful than the ghost controls. The authors concluded that the apparatus manipulations must be performed by a social agent for learning to occur. Ghost controls effectively address the informational equivalence limitation through matching both the information provided in each condition and the required response.

However, differences in reinforcement history between repeating other-generated object motion, compared with spontaneous object motion further impede this design. This is also particularly problematic for between species comparisons. Because animals copy selectively, the behaviour available for copying is typically not random, but biased toward adaptive variants (Rendell et al., 2011). This results from the process of social transmission which streamlines socially available information generating an adaptively-biased subset of behavioural variants (see Rendell et al., 2010). Experience of the higher likelihood of payoff following repetition of others' behaviour plausibly creates a generalised positive association with social copying. Furthermore, this reinforcement history cannot be assumed to be equal between species due to, e.g., differences in species-typical reliance on social learning. Therefore, even when the information content between social and individual learning conditions is matched, defining success by repetition of an experienced behavioural sequence may skew cross-species comparisons. A final issue is that a social agent likely draws greater attention to the content of a demonstration than an invisible one (Fawcett et al., 2002), and again it is possible that this occurs to differing extents for different species.

The current study aimed to minimise the issues discussed by training subjects on the WSLS task where identical informational cues were provided across social and individual learning conditions, and by variably requiring a repeat or explore response depending on the information provided. As the features of the stimuli vary across trials, each exposure to the task represents a novel problem where associations between particular stimuli and rewards cannot be generalised. This allows multiple task exposures with no carry-over effects, and eliminates confounds resulting from differential prior reinforcement histories associated with repetition across the different source conditions. Only through task exposure can subjects learn the WSLS strategy which informs whether a 'stay' or 'shift' response delivers reinforcement on any given problem.

Possibly partly because of the issues described, there are no clear theoretical predictions in the literature regarding how learning and using information from individual sources compares to social sources; however, we can infer what might be predicted. For example, it might be expected that humans would repeat the behaviour of others with more fidelity than non-human animals (henceforth 'animals'), possibly even in circumstances when the behaviour has not directly resulted in a reward, consistent with research on over-imitation (Hoehl et al., 2019). Another possibility would be increased sensitivity to cues associated with reward when exposed to social information. This may result in a greater discrepancy in rates of repetition of reinforced versus unreinforced behaviours when information is acquired socially compared with individually. Indeed, it has been argued that humans are particularly adept selective social learners (Laland, 2004).

In contrast, non-human primates, thought to be poor social learners relative to humans, might be expected to be at a significant disadvantage when learning from the outcomes of others' behaviour compared to their own. Firstly, there is limited evidence of over-imitation in nonhuman primates (e.g., Clay & Tennie, 2018; Horner & Whiten, 2005) and nonhuman animals more generally have been proposed to lack the "explicit" selective social learning strategies of humans (Heyes, 2016). Moreover, examples of culture in nonhuman primates show few signs of directional modification (Tomasello, 1999b). Such evidence has sometimes been interpreted as being congruent with the idea of an adaptation for social learning that evolved late in the hominin lineage (Cosmides & Tooby, 2015). Therefore, it might be predicted that non-human primates would encounter some difficulty, especially relative to humans, in learning the WSLS strategy from a social source.

However, these indicators of apparently poor social learning, relative to individual learning, might be attributable to general cognitive processes, such as the fact that learning from the outcomes of another's behaviour requires active attention to be paid to the model during the critical moment(s) of the task response. Indeed, monkeys are documented to have poorer executive control of attention than humans (Beran et al., 2016). In contrast, when individually learning, this level of attention is necessarily already enlisted by virtue of direct engagement. As such, absence of an evolved social learning module may not be necessary to explain difficulties in basic social learning propensities in monkeys. If attending to social information is more difficult, or less likely, than it is for information generated individually, we would also predict slower mastery of the task contingencies for those exposed to social information. However, it would also be expected that any such difference between social and individual information conditions would reduce as subjects became task competent (having learned through experience to attend to cues in the social condition).

The WSLS task enables us to disentangle these influences by examining rates of repetition when learning from each *information source* (social vs individual) and *information type* (rewarded vs unrewarded). The first stage of the current experiment (Stage A, 2-stimuli) involved training the subjects on the WSLS task which requires flexible discrimination between cues associated with the presence or absence of rewards. This allowed rates of learning the WSLS task rules to be compared between individual and social learning contexts. Any advantage of learning individually at Stage A could be congruent with an inherent difficulty with using socially acquired information, although as noted above, attentional factors could also potentially account for this finding.

Once the strategy was used consistently enough to meet our performance criterion on Stage A, this signified that the informational cues were reliably attended to. The ability of such proficient monkeys to apply the rules learned during Stage A to Stage B (3-stimuli) then tested the fidelity of information use between stages. ‘High fidelity’ in this context refers to high levels of accuracy in the use of information (e.g., evidence of precise repetition of rewarded selections). Any source biases shown by proficient monkeys who used the WSLS strategy with high fidelity upon transfer to Stage B (where sufficient attention could be assumed due to reaching criterion performance) would provide strong evidence that the source of information itself influenced the variation. Alternatively, comparable information use between sources, in combination with initial differences in rates of learning, would suggest no fundamental difference for learning in the social domain, but might point to other source-related disparities, such as differences in attention.

Only one of the squirrel monkeys achieved proficiency on the WSLS in Renner et al. (2019), meaning these hypotheses could not be adequately tested. However, a generalised tendency to repeat was found regardless of information source or information type and it was suggested to represent a pre-existing bias.

Capuchin monkeys may be a better candidate for acquiring the WSLS strategy, from both individually generated and socially observed cues, as they possess several traits suggestive of capacities for complex culture. For example, there is evidence of traditions in wild groups (Perry, 2011; Perry et al., 2003) and they are proficient tool users (Boinski et al., 2000). Capuchins possess allometrically large brains and have a tolerant social system (Perry & Rose, 1994), traits thought to be linked to propensities for culture. In addition, capuchins have displayed the ability to copy with relatively high fidelity (Custance et al., 1999; Dindo et al., 2008; Fredman & Whiten, 2008; O’Sullivan et al., 2017) even without rewards (Bonnie & de Waal, 2007). Of particular relevance to the current chapter, young capuchins appear to prefer to observe proficient nut-cracking models in the wild, suggesting that wild capuchins can learn to discriminate successful behaviour in others to some degree (Ottoni et al., 2005).

To further understand the paradox that many species demonstrate culture but not cumulative culture, the ability for non-human primates to use social relative to individually acquired information warrants further investigation. Any fundamental difficulties found on the WSLS task, particularly if these are restricted to cases where information has been accessed vicariously, may inform our understanding of cognitive barriers to cumulative culture in animals.

Predictions

On Stage A, we expected performance to improve over sessions, consistent with learning of the task contingencies. We planned to compare rates of learning between the group trained on cues provided by social demonstration, and the group trained on cues generated by their own responses.

We also planned to look for any performance differences between the two groups in relation to whether the information trial was rewarded or unrewarded.

On Stage B, we compared whether information source influenced generalisation of the WSLS strategy by individuals that had met criterion on the two-stimuli stage.

Stage A Method

This experiment was pre-registered at the Open Science Framework (osf.io/9f26j).

Subjects & Site

Data was collected from two groups of tufted capuchin monkeys, totalling 35 individuals, housed in separate enclosures at the Living Links to Human Evolution Research Centre at RZSS Edinburgh Zoo. Each group was housed together with a group of squirrel monkeys (*Saimiri sciureus*).

The enclosures were near-identical where each group had access to one indoor (7 x 4.5 x 6m height) and one outdoor (app. 900m^2) area, and the research room (during research sessions) where all training and testing was performed. Both indoor and outdoor enclosures were equipped with a variety of enrichment such as grass, trees, shrubbery, large climbing structures and wood chip. Subjects received two main feeds and three to four scatter feeds daily of a mixture of fresh fruit, monkey chow, boiled eggs, meat, insects and vitamin supplements. Water was available ad libitum. See Leonardi, Buchanan-Smith, Dufour, MacDonald, & Whiten (2010) for full details.

Fourteen subjects took part in the minimum number of sessions (see '*2.1.4 Procedure*') to be included in the analysis (mean age at start of the experiment = 8 years, SD = 4.65; seven females). Subjects were never food or water deprived and all experimental procedures were voluntary.

Ethical approval was obtained from the Animal Welfare and Ethical Review Body (AWERB) of the University of Stirling, ref: (16 17) 74. The research was approved by the Living Links to Human Evolution Research Centre at RZSS Edinburgh Zoo.

Materials

Subjects were tested in a block of eight Perspex research cubicles arranged in a 2 x 4 matrix. Each cubicle measured 49.5 x 52.1 x 51.4cm, and cubicles were separated by partitioning slides during testing. The wall of each cubicle facing the screen incorporated two holes: one enabled the subject to freely reach through the Perspex to touch all areas of the screen, the other to receive food rewards.

The experimental materials were created using PsychoPy 1.83 (Peirce et al., 2019), and were run on a Microsoft Surface tablet computer connected to an ELO 1939L touchscreen monitor. The stimuli were geometric shapes of varying colours and sizes generated by PsychoPy. The monitor was mounted on a portable trolley for presentation to the subjects. An Apeman camera was used to video record sessions and was attached to the inside of the cubicles in a polycarbonate box. Raisins and sunflower seeds were given as food rewards.

Procedure

Subjects entered the cubicles and were voluntarily isolated using sliding partitions. The subjects had previously been trained to signal to end a session and leave the research cubicles at any given point by touching one of the sliding doors. A sunflower seed was given to subjects for isolating in a cubicle, and another on conclusion of the session. Subjects were presented with the ELO touchscreen by pushing the trolley against the front cubicle wall.

Pre-Task Touchscreen Training

As subjects were naïve to touchscreen computers they were initially trained to interact with the apparatus. This training used shaping and a stimulus fading procedure as in Renner et al. (2019). Briefly, this involved 3 consecutive stages where subjects were rewarded with a food item (raisin) and an auditory 'click' for i) touching anywhere on the screen, ii) touching large stimuli on the screen and iii) touching small stimuli (the size used in the task described below). The same arbitrary visual cue (Figure 2.1) as would be used in the experimental task replaced the stimulus on stages ii) and iii) once the stimulus had been selected. Each session consisted of ten trials. Subjects were required to complete 80% of trials per session (8/10) across three consecutive sessions to pass each training Stage and progress to the experimental task (pre-registered). Each trial had a time limit of 30 seconds.

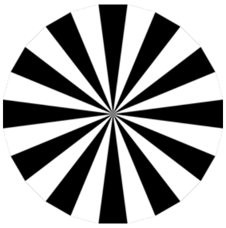


Figure 2.1. Arbitrary visual reward cue used during training and task.

Task Procedure

The procedure was identical to Renner et al. (2019) but will be outlined here. There were four problems per session and each problem included one information trial followed by four test trials. For the information trial, the experimenter (social condition) or the subject (individual condition) made a choice between two horizontally aligned stimuli displayed on the screen (Figure 2.2). Stimuli colour and shape (diamond, square, hexagonal) as well as background colour were randomly generated. Colours were within the range known to be discriminable by capuchin monkeys (Waitt & Buchanan-Smith, 2006).

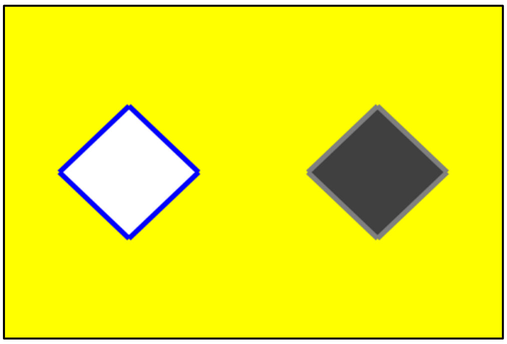


Figure 2.2. Stage A example stimuli.

Within a session, two problems involved a rewarded and two involved an unrewarded information trial. Their order was randomly allocated by the program whereby the stimulus selected by experimenter or monkey was automatically assigned to the randomly generated information type. In the social condition the experimenter used the function *sample* on RStudio (RStudio Team, 2014) to randomly order which stimulus would be chosen (left or right). Following a rewarded information trial, the visual reinforcer (Figure 2.1) replaced the stimulus on the screen and the 'click' sound was simultaneously produced; subjects in the individual condition also concurrently received a food item (one raisin). Selection of an unrewarded stimulus produced no reinforcers and initiated a timeout of three seconds during which the stimuli disappeared and the background colour only was displayed.

For each problem, four test trials followed whereby subjects were presented with the same stimuli as the IT and made one selection per trial. The location of the rewarded stimulus was consistent for each trial within a problem. Only the first test trial was included in the analyses here; test trials 2-4 were included to scaffold the learning of the predictive relationship between the information trial and test trial, and to be used as part of our performance criteria (see below). Following a rewarded information trial, subjects received all three reward cues (visual, auditory and food) for selecting the same stimulus as the information trial (win-stay) and were not rewarded for choosing an alternative (win-shift). After an unrewarded information trial, subjects were rewarded for choosing the alternative stimulus (lose-shift) and not rewarded for choosing the same stimulus (lose-stay). If subjects made four consecutive unrewarded selections, there was a brief time out and a sunflower seed was given to maintain motivation. Subjects had to complete at least one full problem (information trial and all four test trials) for the session to be included in the analysis (pre-registered).

Subjects met criterion on the task when they achieved ≥ 75% correct (i.e., repeating rewarded, and avoided unrewarded stimuli) on both test trial 1, and the aggregate of test trials 2-4 over three consecutive sessions (pre-registered). This indicated they had acquired the successful strategy, and allowed progression to Stage B (see '*Transfer*' sections). If criterion was not met within 30 sessions, testing ceased for that individual (pre-registered). To be included in the analysis, subjects had to take part in a minimum of ten sessions (pre-registered).

There were two opportunities for testing the monkeys per day (am and pm research blocks). Subjects were not tested more than once per block and thus took part in maximum two sessions per day.

All data collection took place from May 2017 to January 2018.

Analysis

Two logistic generalised linear mixed models (GLMMs) were constructed to test our predictions regarding learning on Stage A, largely corresponding to those by Renner et al. (2019).

For model one (pre-registered), task success on test trial 1 was the outcome variable. Responses were recorded as successful if subjects performed a win-stay or lose-shift response, and unsuccessful if they performed a win-shift or lose-stay. Information source (social vs individual), information type (rewarded vs unrewarded), session number (as a proxy for experience) and all their interactions were included as fixed effects. Subject ID and location of reward (left or right) were added as a random intercepts due to the repeated measures nature of the study, and to control for location biases, respectively. Information type was included as a by-subject random slope effect to allow the gradients for each subject to vary.

If the choice made during the information trial was repeated on test trial 1, we classified this as 'repetition', and choosing an alternative was a 'shift' response. Repetition was the outcome variable for model two (not pre-registered) and had the same fixed and random effects as model one, except session number was omitted. This model was constructed to ascertain whether the subjects simply performed the same response regardless of the information trial.

All models had maximal random effects structures (Barr, 2013). Where singular fit or convergence issues were encountered, we adopted a protocol whereby random slopes were initially removed to address this, followed by random intercepts if necessary; however no such issues were encountered.

Responses on test trial 1 were recorded as binary for each outcome variable (successful [1] vs unsuccessful [0]; repeat [1] vs shift [0]) thus, the models were conducted with a binomial distribution function. To run the models, the binary fixed effects (information source and type) were sum coded so that their levels were -1 and 1. Session number was also centred to account for variation in number of sessions completed by subjects.

A Kaplan-Meier survival analysis and Peto-Peto weighted log-rank test were also performed to determine whether information source (social vs. individual) influenced the likelihood of meeting criterion. The survival object included whether the subject met criterion or not as the event, number of sessions as the time variable and information source as a predictor variable. The Peto-Peto modification was used as this removes the assumption of proportional hazards and assigns slightly larger weights to earlier events (Karadeniz & Ercan, 2017). The latter was appropriate as task experience, which increased with exposure, made the probability of meeting criterion less likely at earlier time points.

Models were created using the *glmer* function of the *lme4* package (Bates et al., 2014), the survival analysis used the *survival* packages, and graphs were created using *ggplot2* and *survminer*, all performed on R (R Core Team, 2020).

Stage A Results

Nine of the fourteen subjects met our performance criterion indicating that capuchin monkeys could learn to use the information provided to guide their behaviour. Figure 2.3 displays the number of sessions to reach criterion for all subjects, and indicates source condition: individual (the subject itself performed the information trial), or social (the experimenter performed the information trial). Table 2.1 indicates the proportion of success and repetition on test trial 1 for all subjects, un-averaged.

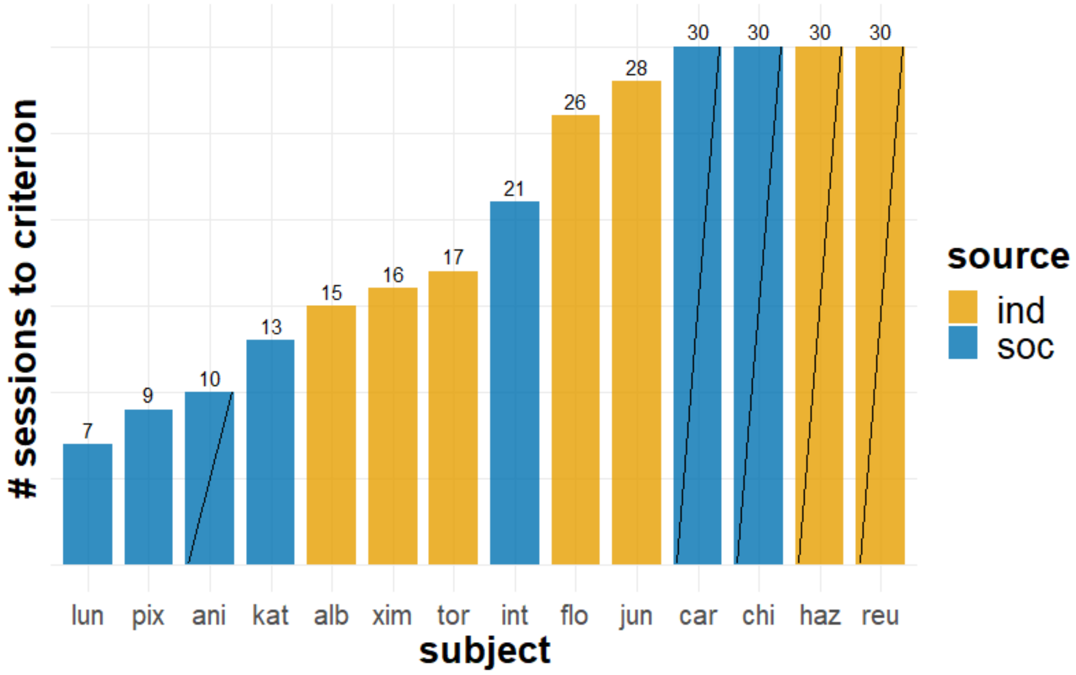


Figure 2.3. Number of sessions to reach criterion and condition of information source received. *Diagonal line through indicates subjects who did not meet criterion.*

Table 2.1. Proportion of WSLS strategy success and repetition for all subjects across all trials, un-averaged. *Chance level is 0.50.*

		WSLS success	***Repetition***
Information Source	*individual*	0.56	0.71
	social	0.59	0.63
Information Type	*rewarded*	0.74	0.74
	unrewarded	0.39	0.61

WSLS Success GLMMs

The GLMM was significantly better than its null equivalent (χ^2 (7) = 38.57, *p* < .001). Overall, increased success was found with exposure to the task (session number; *b* = 0.02, SE = 0.009, z = 2.22, *p* = .03) indicating that performance improved with experience, which appears to be driven by subjects that met criterion (Figure 2.4).

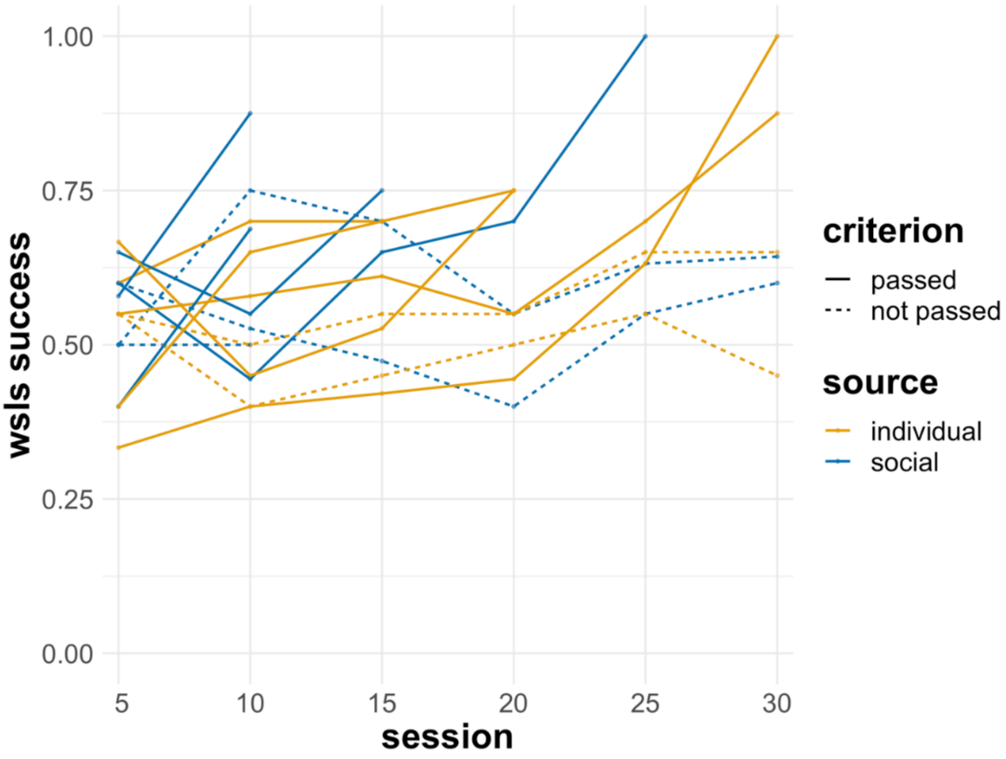


Figure 2.4. Success over sessions separated by subject, whether criterion was met or not and information source; sessions grouped into blocks of five. *NB: the three sessions where criterion was met are included; testing ceased on this stage after criterion was met.*

Significantly greater success on rewarded compared to unrewarded problems was found (*b* = 0.76, SE = 0.09, z = 8.77, *p* < .001) (Figure 2.5) indicating that subjects performed better when a win-stay response was required, compared with lose-shift. Figure 2.6 displays the first and last five sessions on Stage A by subjects that met criterion only. It indicates strong performance on rewarded problems from the outset, whilst responses to unrewarded problems were initially poorer, but generally appeared to improve by the conclusion of this stage.

Information source was not a significant predictor (p = .55) (Figure 2.5), suggesting no differential effect depending on whether the information was generated by the experimenter or the monkey itself. However, it is also possible that limited power may have hindered detection of an effect.

Finally, the interaction between information source and type was marginally significant (*p* = .055). The remaining interactions were not significant (*p* ≥ .31) (Figure 2.5). Thus, the source of information also did not appear to influence learning on this task, or responses to each information type.

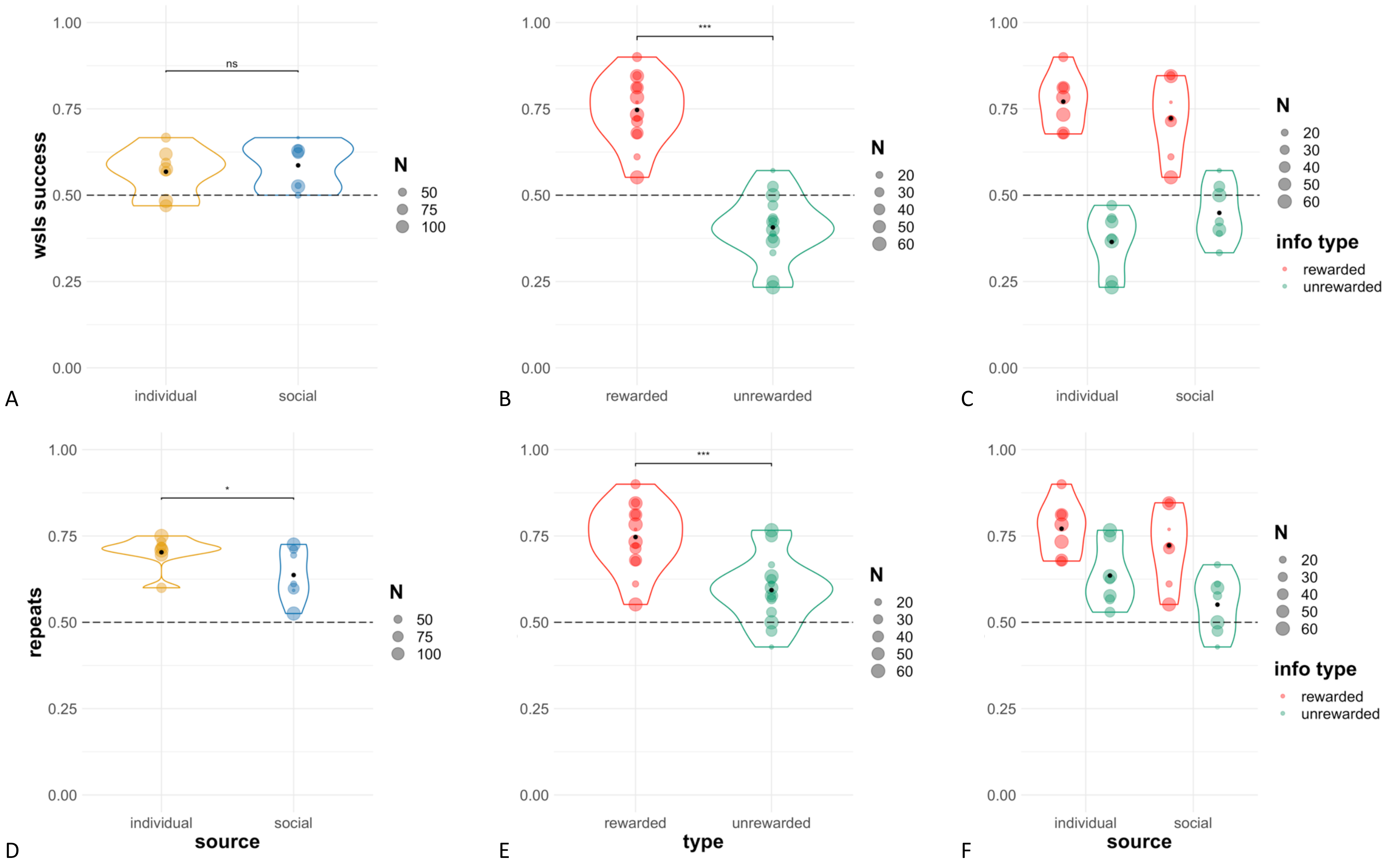


Figure 2.5. Success (top) and repeats (bottom) on Stage A for all subjects by information source (left), information type (middle), and the interaction between information source and type (right). *For rewarded information types, the successful strategy on both measures (WSLS and repeats) resulted in an increase, whereas for unrewarded problems this resulted in an increase in WSLS success, but a decrease in repetition. Dashed line indicates chance; black point indicates mean. 'N' indicates the number of data points (trials) for each individual subject.*

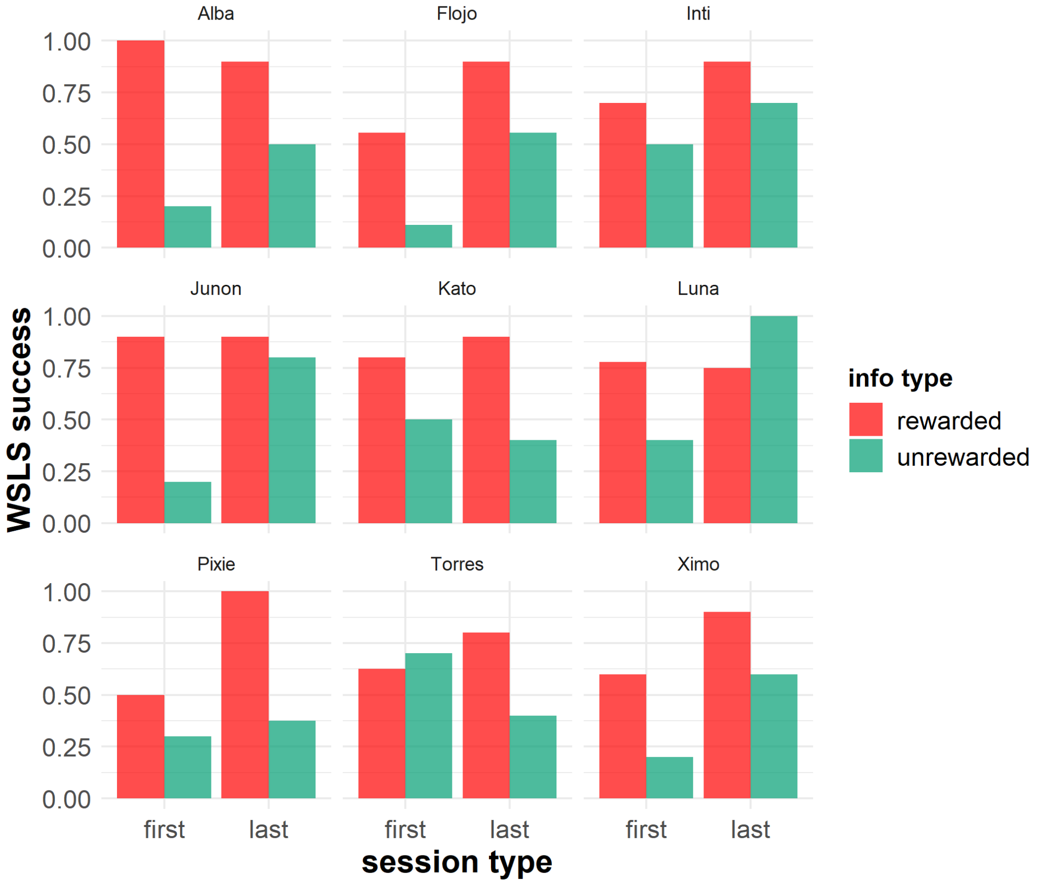


*Figure 2.6. **Use of WSLS strategy by subjects that met criterion, separated by information type.** The first and last five sessions only are included as a broad overview to each subject's progress on Stage A.*

Repeats GLMMs

The repetition model was significantly better than the null model (χ^2 (3) = 16.33, *p* < .001). Information source was a significant predictor whereby the individual condition (*b* = 0.17, SE = 0.08, z = 2.05, *p* = .04) (Figure 2.5) was associated with increased repeats. Thus, subjects who were repeating their own selections were marginally more likely to do so compared with the group repeating selections made by the demonstrator. Consistent with the reinforcement contingencies of the task, the rewarded information type was associated with significantly more repeats than unrewarded (*b* = 0.35, SE = 0.08, z = 4.16, *p* < .001) (Figure 2.5).

The interaction between source and information type was not significant (*p* = .60) (Figure 2.5).

Survival Analysis

Finally, a Kaplan-Meier survival analysis curve was constructed as an additional test of the contribution of information source to the learning process; this perspective assessed whether the social or individual condition was predictive of number of sessions to reach criterion (Figure 2.3). This analysis accounted for the possibly confounding issue that subjects who met criterion at an early stage necessarily contributed less data to the GLMMs described above, than those who required more sessions. Nevertheless, a log-rank test found that information source had no effect on likelihood of reaching criterion; the 'risk' of meeting criterion in the social group was not significantly different to the individual group (χ^2 (1) = 0.6, p = 0.4) (Figure 2.7).

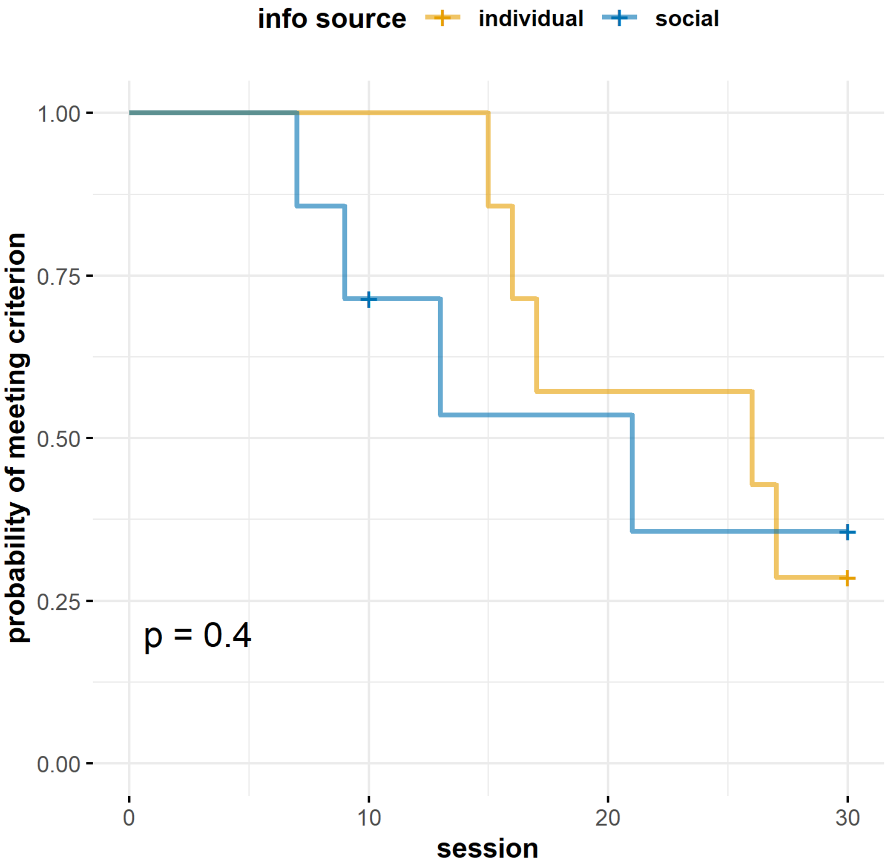


Figure 2.7. Kaplan-Meier curves displaying the likelihood of meeting criterion over sessions for each information source.

Stage A Discussion

Over half of the subjects met criterion indicating that they could strategically use the cues provided in the information trial to some degree, contrasting with the minimal progression of squirrel monkeys on the same task (Renner et al., 2019). However, several subjects did not appear to show improvement, even across the maximum thirty sessions, implying limited awareness of the predictive relationship between the information trial and the test trial.

Learning of the WSLS strategy was not found to significantly vary by whether information was provided through social demonstration or individual experience, despite robust analysis to ensure a potential effect had not been missed. This suggests that because the value of information provided was strictly matched, and repetition was not the only rewarded response, learning to use social information occurred just as readily as learning to use individually acquired information. This finding implies the learning cues were equally potent in this context.

Subjects were found to perform the win-stay response more than the lose-shift; at first glance, this may suggest that the subjects were more proficient at learning the win-stay rule. However, this is unlikely considering that high levels of repetition were found i) across both information trial types, and ii) following minimal exposure to the task. These points instead indicate that 'stay' was simply the dominant response overall, irrespective of the learning that occurred as a result of task experience.

Thus, an explanation that better fits our pattern of results is that the monkeys had a pre-existing generalised bias to repeat, largely regardless of whether this was repetition of their own behaviour or the demonstrator's. This corresponds to the tendency to repeat found by Renner et al. (2019) with squirrel monkeys. The appearance of the reward cue, already reinforced in training sessions, may have broadly heightened interest in the associated stimulus, driving the slightly higher rates of repetition in the rewarded condition. As such, these findings are possible without invoking a strategic understanding of the task contingences, or indeed suggesting that the monkeys treated the IT as information at all. Also similarly to Renner et al.' (2019) 3D version of the task (experiment 2), we found slightly higher rates of repetition in the individual learning condition. Simple behavioural inertia (individual condition) and stimulus enhancement (social condition) likely underlie this repetition bias (Renner et al., 2019).

However, it is plausible that subjects that did meet criterion (henceforth 'criterion subjects') entered the study with this spontaneous bias to repeat but began to treat the information trial as 'information' that they used with an element of cognitive control as the WSLS strategy was acquired. In other words, as the predictive relationship between information trial and test trial was repeatedly experienced, criterion subjects may have begun to copy rewarded and inhibit repetition of unrewarded selections, resulting in increased success consistent with a rule being learned.

Importantly, the low levels of learning during most sessions included in the Stage A analysis (i.e., all of the non-criterion monkeys' sessions and the early sessions of the criterion subjects) may have obscured a potential effect of information source, which was predicted to occur when learning was actually taking place. As learning presumably occurred mainly in the sessions immediately before criterion was met, any influence of information source may be expected mainly during those sessions.

To deepen our understanding of the influence of information source on strategic information use by criterion subjects, they were presented with a three-stimulus version of the task in a second stage (B); again, only one stimulus was rewarded (Figure 2.8). Increasing the number of distractor stimuli reduced the likelihood of repetition by chance; the ability to generalise the WSLS strategy was then evaluated by comparing performance from Stages A and B.

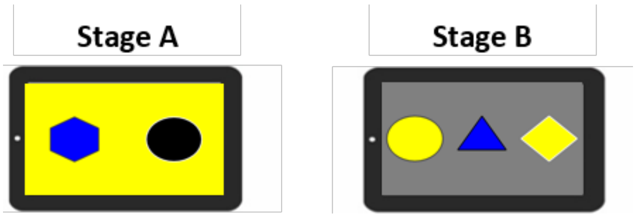


Figure 2.8. Example stimuli from each stage. *One selection was made, and one stimulus rewarded at both stages.*

Although we found minimal effects of source in Stage A, we were interested in whether the source of information may affect the fidelity of information use on transferral to a new context. If information were used with high fidelity, we might expect for example that rewarded stimuli would be repeated at similar rates regardless of the number of distractor stimuli. Such responding would suggest that the WSLS rules had been learned. However, if rates of repetition of rewarded stimuli were lower in Stage B relative to Stage A, this would be more consistent with low fidelity use of information. Essentially, the question is, have the monkeys learned *to repeat* (or to avoid) one particular selection through their prior exposure to its reward value (in which case we would expect to see this applied with high precision)? Or alternatively, does the information trial simply induce a bias towards a particular selection, such that rates of repetition will be strongly affected by the baseline probability of making that selection relative to alternatives.

Because most sessions included in Stage A occurred before any of the WSLS strategy was acquired, i.e., before criterion had been met, we considered a subset of the criterion monkey's data for the transfer analysis (see Analysis below) to mitigate the potentially 'noisy'/diluting influence of pre-acquisition sessions. This analysis may constitute a more valid test of the influence of information source by considering only the most proficient monkeys who had presumably engaged in at least some learning to attend to the task cues. Although this translates to a small dataset, we believe these are worthwhile analyses due to their central role in discussions of comparative cumulative culture.

As the information provided on unrewarded trials did not appear to have been used with proficiency on Stage A, any improvement in response to unrewarded information was also of interest. As the successful response to unrewarded trials is incongruent with the monkeys' predisposition to repeat found on Stage A, the bias may take significant effort to overcome.

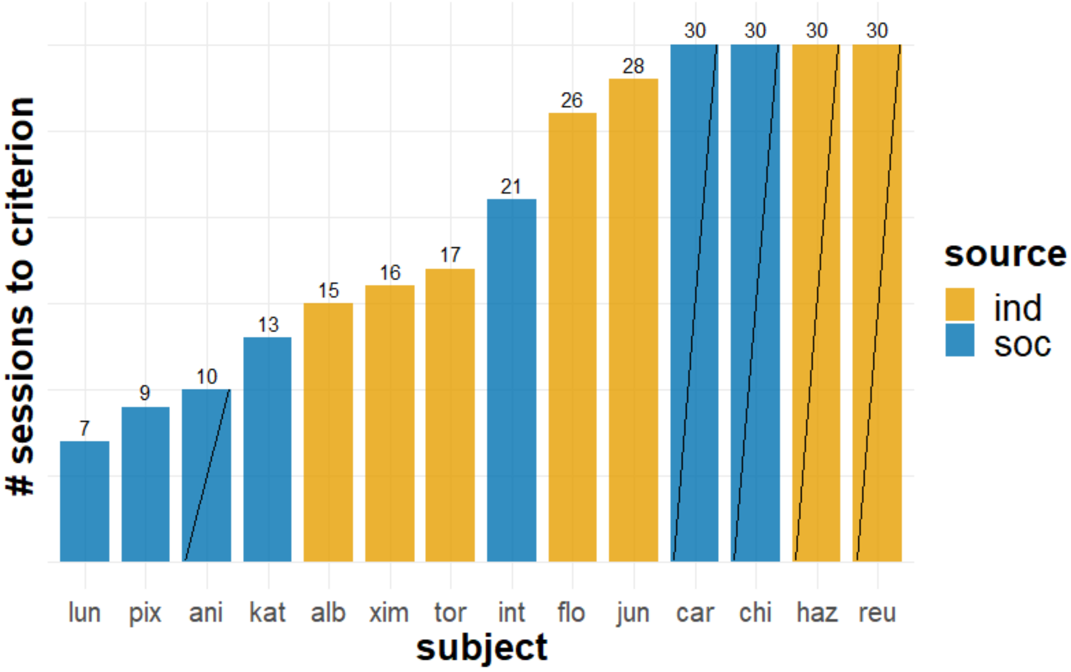
In the following section we examined whether information source affected generalisation of the WSLS strategy on transfer from two to three stimulus arrays. Any consistent differences found between the source conditions here would suggest that processing of information is fundamentally affected by the origin of that information.

Transfer Method

Stage B was identical to Stage A except for some key details outlined below.

Subjects

Nine monkeys reached our pre-determined performance criterion on the two-stimulus task (Figure **2.3**); seven out of these nine (five from the individual and two from the social condition, the criterion subjects) were transferred to Stage B to assess their ability to generalise the WSLS strategy. For the two of the original nine subjects that had met criterion but were not transferred, there was a gap of more than two weeks between reaching criterion for Stage A and the beginning of Stage B. Thus, they were required to meet criterion again over two sessions on Stage A to be transferred but did not achieve this. Since the intention was to test only individuals who had achieved proficiency in Stage A so that we could determine how this learning was generalised, these individuals were not included in the Stage B testing.



Procedure

The key task alteration was that the number of stimuli available to choose from increased from two to three. As in Stage A, one stimulus was rewarded (the other two were unrewarded), and two information trial selections per session were rewarded and two were unrewarded. However, one unrewarded problem per session during Stage B required subjects to choose an alternative stimulus twice to locate the target; use of the correct strategy on test trial 1 (i.e., lose-shift) regardless of finding the target was recorded as a successful trial. As test trial 1 only was included in the analysis, the outcome of the subsequent test trials did not affect the results.

Analysis

A logistic GLMM was performed on the Stage B data only to test the influence of information source and type on the 3-stimuli stage. As with the Stage A GLMMs, data were included for all criterion monkeys up to their thirtieth session for Stage B. The model included repeats on test trial 1 as the outcome variable, and information source, information type and their interaction as fixed effects. Subject ID was included as a random intercept and information type was included as a random slope. All fixed effects were transformed by the same processes as in the Stage A GLMMs. Information type was dropped as a random effect on model two due to non-convergence, but all other models were maximal.

The last three sessions before criterion was met on Stage A, and the first three sessions of Stage B were included in the following analyses. This allowed comparison between conditions once subjects had reached a suitable level of competence with the task. The three sessions where subjects met criterion were not included in any analysis as this would necessarily include sessions where success of 75% or above was achieved, but are graphed below for illustration only.

Four exact binomial tests were performed to test whether proportion of repeats on test trial 1 was significantly different to chance for each combination of levels: i) Stage A, rewarded ii) Stage B, rewarded iii) Stage A, unrewarded and iv) Stage B, unrewarded.

Two chi-square tests (one for each information type) were also performed to determine any statistical differences between proportion of repetition across stage (A vs B).

The binomials were conducted using the *binom.test* function and the chi-squares using *chisq.test*, all on R (R Core Team, 2020).

Transfer Results

Repeats GLMM

The GLMM was significantly more explanatory than the null model (χ^2 (3) = 150.25, *p* < .001). The only significant effect was of information type (*b* = 1.30, SE = 0.12, z = 10.62, *p* < .001), as there were significantly more repeats following rewarded compared to the unrewarded trials (Figure 2.9), in line with the correct strategy. While information source (*p* = .47) did not have a significant effect on repetition (Figure 2.9), the interaction between type and source was non-significant, although close to our alpha criterion (*p* = .051).

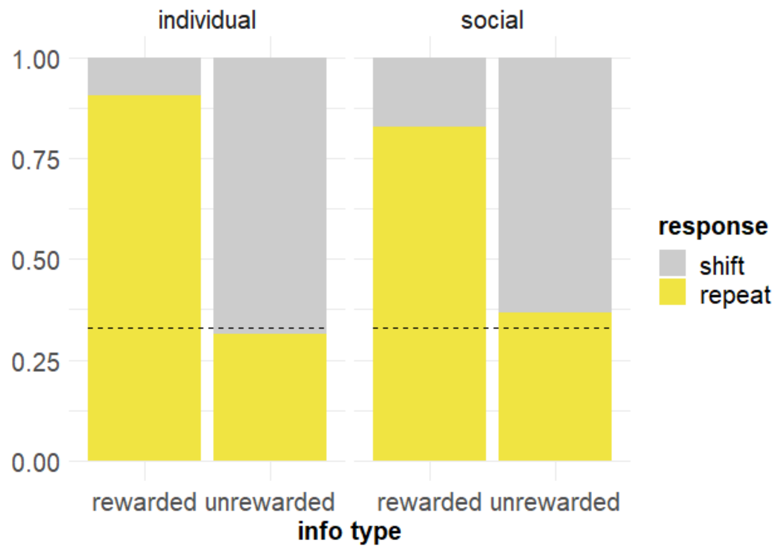


Figure 2.9. Repeats on all sessions of Stage B broken down by information type and faceted by information source. *The optimal response for rewarded trials was to repeat, and for unrewarded was to shift. Dashed line indicates chance.*

Binomial and Chi-Square Tests

All results in the remainder of this section will be considered relative to the associated chance of repetition (Stage A = 0.50; Stage B = 0.33).

Table 2.2 displays the proportion of repetition at each stage[1].

---

[1] The breakdown in the number of trials included in each analysis in this section are as follows: three sessions per stage (A and B), two problems per information type (rewarded vs unrewarded), for seven monkeys (3 * 2 * 7) = 42 trials per information type, per stage. One monkey (Flojo) from the individual group did not complete one trial per stage, meaning there were 41 trials per stage for the individual condition.

Table 2.2. Proportion of use of repeats, split by information source. *The three sessions prior to meeting criterion on Stage A and the first three sessions of Stage B are reported.*

			rewarded	***unrewarded***
Repeats	**Stage A (pre-crit)**	**social**	.67	.58
		individual	.83	.72
	Stage B	**social**	.92	.42
		individual	.77	.34

Rewarded Problems

Two exact binomial tests found that the proportions of rewarded information trials repeated during Stage A (.79; $p < .001$) and Stage B (.81; $p < .001$) were significantly greater than the expected chance proportions of .50 and .33, respectively (Figure 2.10). Furthermore, a chi-square test of independence confirmed that there was also no statistical difference between proportion of repeats across the stages (χ^2 (1) = 0.07, $p = .79$). Thus, the drop in chance of repetition from introducing more distractors did not result in any decrease in rates of repetition, suggesting the information was used with high fidelity.

Unrewarded problems

The proportion of unrewarded information trial selections repeated during Stage A (.68; $p = .03$) was also greater than by chance, however by Stage B (.37; $p = .74$) subjects were not significantly different to chance (Figure 2.10). A chi-square test of independence found a difference in the proportions of repeats across Stage A and B (χ^2 (1) = 8.26, $p = .004$). These results suggest that on acquiring the WSLS strategy on Stage A (i.e., meeting criterion) and progressing to Stage B, capuchins' repetition of unrewarded demonstrations moved from greater than chance, to chance levels.

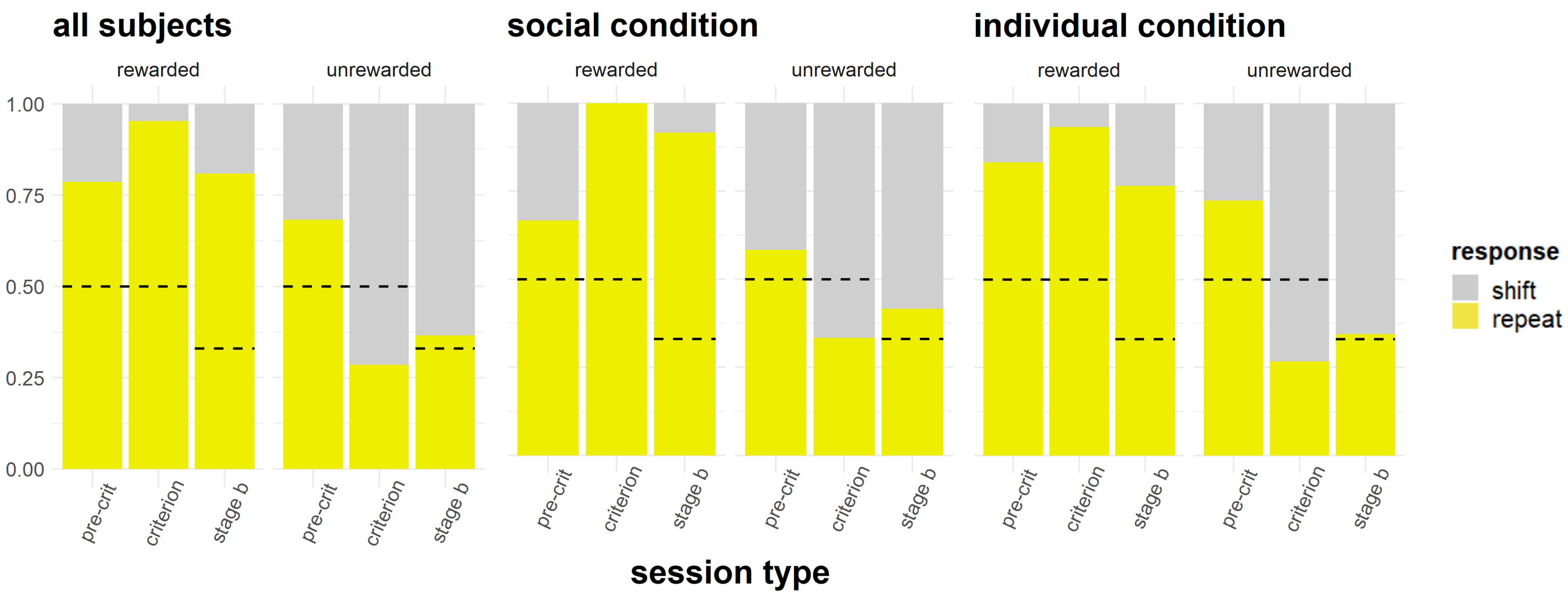


Figure 2.10. Response (repeat or shift) to rewarded and unrewarded information trials by all criterion subjects (left) and separated by information source (middle and right). *The last three sessions of Stage A before criterion was met ("pre-crit"), three sessions where criterion was met on Stage A ("criterion", included for illustration only) and first 3 sessions of Stage B ("Stage B") are shown. The optimal response for rewarded trials was to repeat, and for unrewarded was to shift. Dashed line indicates chance (Stage A = .50; Stage B = .33).*

Transfer Discussion

We found no evidence that there was any fundamental difference in the fidelity with which information was used between the individual and social learning conditions when the number of distractor stimuli was increased, even considering only subjects of higher task proficiency. This is similar to the results of Renner et al. (2021) who found little influence of source condition on proficient monkeys in a within-subjects variation of the current task (carried out after the current study).

Our results do suggest that our criterion subjects had learned the win-stay rule as their ability to successfully respond to the reward cue (by repeating) was found to be robust across contexts, despite the lower probability of doing so on the Stage B of this study. In addition to simply being significantly above chance, a similar level of repetition was maintained on progression to Stage B, implying minimal variation in rates of copying was generated by the introduction of another distractor. As such, this species does appear capable of learning and generalising a simple rule-based strategy.

As subjects did not master the lose-shift response we cannot fully test the generalisation hypothesis in this condition. Learning the required response for unrewarded information trials may have been more difficult for this species as it involves inhibition of a pre-existing tendency to repeat (as previously discussed). This contrasts with rewarded problems where this tendency was congruent with the optimum strategy, which was therefore frequently positively reinforced. Although the subjects did not approach proficiency on unrewarded problems, there may have been a trend toward decreased repetition (from significantly above chance to non-significantly different from chance) on transferral to Stage B. Of course, this is somewhat conjectural given the small quantity of data included in this analysis. However, potentially, as our subjects became selectively more attentive to stimuli associated with the reward cue, they became less attracted to unrewarded stimuli which were not linked with this cue. The chance level performance on unrewarded problems in Stage B may indicate that unrewarded information trial selections stimulated no particular interest in any of the available stimuli, whereby the unrewarded cue was still not treated as 'information'. This relatively neutral response may be a precursor to actively avoiding the unrewarded stimuli, and thus engaging in the optimal lose-shift strategy. This may require increased exposure to the task.

Overall, we found that information about the location of rewards was used with reliability in a fresh context, although the poor performance on unrewarded problems precludes any conclusion regarding the fidelity of use of this type of information. Notably, generalisation of the WSLS strategy was found to have no link to whether the information was provided socially or from individual exploration.

Discussion

Our aim was to train one group of monkeys to learn a WSLS strategy using cues from a social source and another group with cues from individual exploration. The WSLS task is a robust method for comparing these conditions owing to its capacity to control for variation in information delivery between sources and in reinforcement history.

Comparisons of individuals in the social and individual conditions found no evidence that the rewarded or unrewarded cues were learned or used differently, or that information was used with higher fidelity. This suggests that in principle, the study species learned to use social information as effectively as they learned to use information based on feedback from their own activity.

A study testing children recruited in Scotland and China on a homologous version of our task also found that information source had no effect on 2- to 5-year-old children's performance (Atkinson et al., 2020). This similarity between the human and monkey populations implies that, fundamentally, learning from social sources may not be driven by a specialised cognitive adaptation (even in humans) but could be based on general-purpose learning mechanisms (Behrens et al., 2008; Heyes, 2012; Osiurak & Reynaud, 2019). Potentially, at least some of the significant differences found between humans and other animals in typical social learning studies could be attributed to variation in general learning capacities. Of course, there are likely facets of human cognition that are specific to the social domain, but which are unlikely to be adaptations specifically for the acquisition of cultural information, that do play an important role in cumulative cultural evolution. These may be necessary for more challenging tasks than the current one e.g., those requiring the application of information based on theory of mind or meta-cognitive reasoning (Dunstone & Caldwell, 2018; Heyes, 2016).

Our results may contrast with theories which implicate a fundamental difference in the treatment of information gained through direct experience or from social sources, both within and between species. However, other lines of research indicate that our results are not surprising, such as neuroimaging studies reporting that personally executed vs. observed actions activate the same neural pathways (Bonini & Ferrari, 2011). More specifically, this has also been observed when personally experiencing vs. observing negative feedback in humans (Shane et al., 2008; Yu & Zhou, 2006).

The present study did find that the capuchin monkeys struggled to grasp the significance of the task cues, particularly on unrewarded trials. A spontaneous bias toward repetition of behaviour was found, indiscriminately driving high rates of repeat responses to both 'win' and 'lose' problems. It may be that primates have a general difficulty with selective learning, involving sub-optimal adjustment of responses relative to the available information, regardless of the information’s source.

Atkinson et al. (2020), discussed above, found that children successfully adopted both the win-stay and lose-shift rules, using both types of information with high fidelity. However, the children also displayed a striking tendency to explore novel locations (see also Blanco & Sloutsky, 2020), performing far better following *unrewarded* information trials, compared with rewarded ones. These results contrast strongly with the bias toward repetition found in the capuchins here. Furthermore, the children’s performance implies a solid grasp of the task contingencies and flexible information use, whereas the monkeys’ results suggest only limited recognition of the significance of the cues to the reward location.

Despite entering the task with a repetition bias, our criterion subjects appear to have developed some selectivity in their use of the information with task exposure. The generalisation of the win-stay rule indicates that this type of information was used with precision. The trend toward inhibition of lose-stay responses also appears consistent with this interpretation. Developing an understanding that the reward cue resulted in reinforcement for repetition possibly cultivated a strategy such as ‘repeat-when-reward-cue-appears’, which may have rendered the subjects less susceptible to the repetition bias when faced with an unrewarded information trial, and therefore less drawn to the stimulus that was selected.

The shift from reflexive responding to a more strategic approach was likely easier for the win-stay rule, where the repeat bias was congruent with, and positively reinforced by, the task structure. Learning what *not* to do may have been more difficult because the unrewarded cue necessitated a response that conflicted with the demonstration (Brown & Braver, 2005). It may also have been partially due to subjects’ reinforcement history e.g., a pre-existing more powerful association with repeating rewarded behaviour due to the immediate gain, relative to active avoidance of unrewarded behaviour which is less likely to result in instant gratification.

The implementation of the WSLS in the current study involved a degree of memory load for the information trial to be used to guide test trial selections. It was important for attention to be directed both to which stimulus was selected and whether it revealed the reward cue. It is plausible that this cognitive load made the WSLS strategy difficult for the non-criterion monkeys to learn, regardless of information source, and for mastery of the lose-shift rule overall (Wilks et al., 2021). Indeed, working memory capacity has been linked to the emergence of sophisticated culture in humans (Guida, 2020; Wynn & Coolidge, 2007). It has also been documented that non-human primates are more prone to bottom-up, habitual responding than humans who exercise greater executive control of attention (Beran et al., 2016).

Overall, we found an interesting pattern of results that may have implications for understanding why animals display behaviour consistent with cultural learning but little evidence of cumulative culture. The ability to overcome natural biases and approach problems strategically may amount to a significant barrier at the group level where cultural accumulation occurs. In particular, learning to disregard actions that were unrewarded may present cognitive challenges that are too difficult to overcome for non-human primates; for example, storage of negatively-valenced information in order to avoid repeating unsuccessful behaviour in future. In contrast to humans, explicitly strategic, high fidelity use of information may be less available to non-human primates which may significantly contribute to the scarcity of cultural accumulation.

To further understand the factors constraining optimal success on the WSLS, the associated memory load could be removed. Alleviating the cognitive burden of memorising information trial outcomes may increase the capacity for fully-fledged discrimination learning of both types of information, and potentially bolster the success of non-human primates.

If subjects were to develop more reliable responses to the basic WSLS cues, this would permit extension of this paradigm by scaling up both the number of stimuli and the number of selections made per trial. Such a scaling-up would allow investigation of the ability to flexibly use multiple cues of both information types within a single problem. For example, consider a stage where four selections are made from an array of sixteen stimuli with the aim to find 4 rewards (such as the task described in Chapter 1, section '*The PFR Method*'): following an information trial that discovered one rewarded and three unrewarded stimuli, consistent WSLS responding would allow matched or improved performance at the test trial. With repeated exposure to such problems, the optimal strategy would generate an overall increase in success relative to the information trial. Demonstrations of varying value (i.e., the number of reward targets found) could effectively test whether subjects have the potential for cumulative cultural evolution by simulating exposure to information at different stages of transmission (Caldwell et al., 2020). Outperformance of these demonstrations would suggest that the basic capacities necessary for cumulative cultural evolution are available to the study subjects under optimal conditions.

Methodologically, the addition of a post-acquisition phase, after criterion had been met, where subjects continue the same stage for a set number of trials, would represent a more useful test period to observe potential biases or strategies used.

The body of results presented here, by Atkinson et al. (2020) and Renner et al. (2021) suggests that children and capuchin monkeys are, at least in their performance in this context, indifferent to whether information is derived through social or individual learning when the conditions for learning are controlled. Therefore, it is likely that general-purpose learning abilities may at least partially account for differences found between humans and other primates on tests of cultural learning. Future experiments should ensure an adequately matched asocial learning control to test the nature of social information use more comprehensively. However, barriers to general selective learning may constrain the ability for non-human species to repeat useful behavioural variants and intentionally disregard less effective alternatives. Our understanding of why human culture significantly differs from that of other species at a basic level would benefit from a finer grained investigation of the selectivity of learning.

Chapter 3 Validating the Touchscreen 'PFR' Task with Children.

Contributions

Donna Kean, Christine Caldwell, Elizabeth Renner and Mark Atkinson contributed to study conception, design and material preparation. The code for the task was written by Mark Atkinson. Data collection was performed by Donna Kean and Clare Campbell. The data analysis was performed by Donna Kean with contributions from Christine Caldwell. The chapter was written by Donna Kean and edited by Christine Caldwell.

Introduction

The primary goal of this thesis was to test the potential for ratcheting in non-human primates. Ratcheting in this context refers to a measurable improvement in the functionality of a trait resulting from information transfer over generations of learners, in line with definitions of cumulative culture (Mesoudi & Thornton, 2018; Tennie et al., 2009). This likely requires *selective* learning, where information from previous generations is strategically used by subsequent learners, yielding an overall performance increase (Kendal et al., 2018).

Chapter 2 investigated whether a simple selective strategy could be learned by tufted capuchin monkeys (*Sapajus apella*) on a win-stay, lose-shift (WSLS) task, where repetition of rewarded and avoidance of unrewarded behaviours was required. Although mastering this strategy at a basic level would not display the potential for cumulative culture, it may be a fundamental process underlying cumulative culture ability. High-performing monkeys appeared able to learn and generalise repetition of a rewarded selection, whereas correctly responding to information about unrewarded selections appeared difficult for the capuchins to master overall. In a subsequent study on the same test population, however, Renner et al., (2021) found that two monkeys mastered both the win-stay and the lose-shift rules following further training on the same task. Displaying this level of proficiency on the WSLS task is suggestive of some capacity for ratcheting, given enough context-specific experience.

However, to test the potential for ratcheting in any population, there must be a range of possible success scores associated with the task used, allowing the possibility for performance to improve over multiple episodes (through demonstrated solutions being successively outperformed). This is out of the scope of the previously used WSLS task where there were two or three stimuli, only one selection was made per trial, and the only possible demonstration/performance scores were *0* (unsuccessful) or *1* (successful). Performance on the WSLS task also could not simply be extrapolated by assuming that the same trends would continue if, for example, the number of distractor stimuli and the number of selections made on a single trial were increased (creating a range of possible scores). This is because it may be easier to outperform the success of a task-naïve, low-scoring demonstrator than it is to outperform more proficient, higher-scoring demonstrators, who had themselves learned from a previous participant(s) and honed their skills or increased their knowledge. Thus, improving upon higher scoring demonstrations would potentially pose a greater challenge than lower scoring demonstrations, despite the same basic strategy leading to success (repeat rewarded, avoid unrewarded).

Other tasks have previously been successful in displaying cumulative build-up with both humans and non-human animals (henceforth 'animals'). For example, spaghetti towers could reach a wide range of potential heights (Caldwell & Millen, 2008); puzzle boxes could be solved to multiple levels of increasing difficulty (Dean et al., 2012); and the routes of pigeons could vary along a continuum of efficiency (Sasaki & Biro, 2017). Such a range of success scores enables continuing improvement to be observed over learner generations, a process essential to identifying cumulative culture.

To allow for this task feature, we developed an electronic version of a task used to test potential for cumulative culture in young children developed by Wilks et al. (2021). Wilks et al. (2021) is currently the only study that has compared children's capacity for cumulative culture between different age groups, and is based on the method set out by Caldwell et al. (2020). The task they employed was based on the same underlying structure as the WSLS, where rewarded stimulus selections should be repeated, and unrewarded selections avoided. To determine the validity of our new computer-based adaptation of this method, we initially aimed to broadly replicate Wilks et al. (2021) results with a small sample of children. We then aimed to use this method with non-human primate populations to test their capacity for cumulative culture (Chapters 4 and 5).

Wilks et al. (2021) tested 3- to 6-year-old children on a task called 'Find-The-Treasure' that required children to look for three pieces of hidden treasure from 3 x 3 arrays of treasure chests. On each array, a parrot hand-puppet performed an initial demonstration trial where one treasure chest was selected from each row (three selections in total). Chests was either found to contain treasure (rewarded) or was empty (unrewarded). After the puppet had made all three selections, the child also chose three chests from the same array; the locations of the treasure remained constant. Children were instructed to try to find as many pieces of treasure as possible. The optimum strategy was to repeat selections that had revealed treasure and to avoid selecting empty chest selections. Each child completed three trials, each involving a distinct 3 x 3 array of differently coloured chests.

Whilst testing selective information use generally, this stimulus selection task also permitted assessment of potential for ratcheting due to the scope for outperformance of multiple demonstrations. The three trials each child participated in involved demonstrations where the total number of rewarded treasure chests found was either *1*, *2* or *3*. When *1* or *2* treasure pieces were found in the demonstration trial, and if following the optimal strategy, there was potential to achieve higher scores by finding more pieces of treasure (on average). As a result, Wilks et al. (2021) used the children's responses to demonstrations of varying value to infer ratcheting potential for each year banding, in line with the reasoning described by Caldwell et al. (2020). This was done by assessing to what extent they achieved higher scores as demonstration score increased. They were then assigned a position along a continuum of capacity for cumulative culture (discussed further below).

Exposing children to demonstrations of varying value mimicked a core characteristic of generational replacement designs (described in Chapter 1, '*How has the capacity for cumulative culture been operationalised?*') that satisfy the requirements for cumulative culture: later generations, on average, are exposed to demonstrations that contain higher quality information relative to earlier generations. This is an outcome of successive improvement. By inspecting individual level responses in relation to demonstrations of increasing score, Wilks et al. (2021) inferred the outcome of repeated transmission of information. Thus, the experimenters evaluated whether a ratchet effect may theoretically be possible in several age groups.

Although not attempted by Wilks et al. (2021), this method also makes it possible to infer ratcheting at an individual level, as long as the individuals are given multiple exposures to each level of demonstration. This feature of the method makes it particularly advantageous for testing non-human primate populations where access to large numbers is severely limited.

Wilks et al. (2021) task design effectively combatted many issues that have likely contributed to the current dearth of investigations into young children's and non-human's cumulative culture ability. As outlined in Chapter 1, generational replacement studies are widely accepted as the state of the art in assessing potential for cumulative culture, as both the transmission process and the potential for cumulative build-up can be observed in a laboratory setting. These have been successful in studies with human adults (e.g., Caldwell & Millen, 2008; Zwirner & Thornton, 2015). However, as cumulative cultural evolution is a group level phenomenon, a large sample must be obtained to employ these methods, with individual participants strictly organised into sequential chains of learners, and each chain representing a single independent sample unit. If the effects of development are to be studied, then the necessary sample sizes become even bigger as this undertaking must be completed with a range of age groups.

Only a handful of studies have attempted this method with children. For example, Reindl and Tennie (2018) tested 4- to 6-year-old children's ability for cumulative culture using an adaptation of the spaghetti tower task developed by Caldwell and Millen (2008). Each generation of a transmission chain was exposed to the final tower designs of the two preceding generations, allowing the potential to learn from end-state products and build taller towers. Across eight chains of ten generations each, the children displayed no benefit of observing the artefacts of earlier generations, as no cumulative improvement in success relative to chain position was identified; subjects later in the chains were no more successful than earlier generations, nor indeed baseline conditions. Overall, the authors suggested that children of this age range did not display potential for ratcheting. However, age effects were not addressed in this study as multiple replicates for individual age groupings were not attempted here. Therefore, it is unclear whether cumulative build-up may have been possible in certain age groups only.

As discussed in Chapter 1, the PFR (i.e., potential-for-ratcheting) Method can mitigate some of the practical issues associated with typical transmission chain procedures by inferring potential for ratcheting at the individual level. The success of its first implementation by Wilks et al. (2021) provided validation of the method's logic. The current project ultimately aimed to use this logic to test non-human primate potential for ratcheting using an electronic homologue of Find-the-Treasure. Initially however, the current chapter sought to test its efficacy with children as this allowed validation against the results of Wilks et al. (2021).

An additional manipulation by Wilks et al. (2021) demonstrated the importance of the context of learning when assessing ratcheting between categories of participants. During test trials, cues regarding the treasure chests chosen during the puppet demonstration either remained visible or were hidden from view. The *visible cues* condition minimised memory load by providing a hint regarding whether treasure was found (or not) on each demonstration trial selection; the treasure chests chosen were left ajar, so that whether they contained treasure, or were empty, was partially visible while the children made their selections. In the *transient cues* condition, the demonstration itself took the same format, however the hints were available for a short time only as the chests were closed after the demonstration, so that their contents were not visible during test trials. This required participants to store both which chests had been selected by the puppet and whether they were rewarded or not.

According to Caldwell et al. (2020), for a group (e.g., the age groups used by Wilks et al., 2021) or an individual to be classified as having the capacity for ratcheting when employing this approach, two main criteria are required. Firstly, the individual or group must display above chance performance overall, which demonstrates that they utilised the information available to perform better than without such exposure. Secondly, success must be linked to demonstration value, such that higher performance follows higher scoring demonstrations. These processes are fundamental to cultural accumulation. Outperformance of multiple levels of demonstration would further indicate that ratcheting may be possible in a linear transmission scenario.

Whereas all year bandings displayed the potential for cumulative build-up when cues remained visible in Wilks et al. (2021), only the oldest age group (age 6) demonstrated capacity for ratcheting in the transient cues condition. These results were in line with the authors' predictions that increasing the cognitive demands of the task (i.e., memory load) would increase the age at which potential for ratcheting would be found. Clearly, ratcheting potential between different age groups was dependent on the contextual features of the task.

This is particularly relevant from the perspective of the current thesis, as the previous chapter involved 'transient' cues, in that the demonstration trial cues disappeared before the monkeys made their own selections in the test trials. Since that task (involving just single selections from only two/three different stimuli per trial) proved very challenging for the monkeys, Wilks et al. (2021) results suggests a potentially valuable alternative approach, which might yet allow potential for cumulative culture to be evaluated in nonhuman primates. It is plausible that under optimal conditions, e.g., no associated memory load, increased propensity for cumulative culture may be found compared to tasks with a higher cognitive burden. That is not to say that subjects may advance from no ratcheting potential whatsoever to full blown cumulative culture, but, more likely, they may simply display abilities that would rank higher on a spectrum of ratcheting capacity.

To test non-human primate capacity for cumulative culture we developed a method that followed the reasoning of Caldwell et al. (2020). The current chapter first validated this method against an existing study with children that used a logically equivalent task (Wilks et al., 2021). It was important to demonstrate that the new task could identify potential for ratcheting within a population that had already demonstrated this capability. This would allow greater confidence, whether a ratchet effect is found or not in non-human primate populations, that this is a reflection of their cultural capacities and not some methodological issue introduced by the new task design or presentation medium. The following section broadly outlines how the method described by Caldwell et al. (2020) was operationalised for the current chapter, and with non-human primates in Chapters 4 and 5.

The PFR Method

The PFR Method involves exposure to the PFR Task, followed by assessment of potential for cumulative culture using both the PFR Classification System and transmission chain simulations.

The PFR Task is a stimulus selection task that requires selective information use in order to find as many targets as possible. The touchscreen PFR Task used in this and the subsequent chapter has four stages of increasing complexity (Table 3.1) where the number of stimuli, the number of targets available to find, and the number of selections made (by the demonstrator and participant) are steadily increased. Only at Stages C and D is there a sufficient range of possible reward outcomes for ratcheting to be appropriately assessed, as these stages allow a sufficiently broad range of possible task scores.

Table 3.1. The PFR Task stages.

Stage	***Number of stimuli***	***Number of targets, and selections made***
A	2	1
B	4	2
C	9	3
D	16	4

Stage C corresponds relatively closely to the 3 x 3 grids used in the Find-the-Treasure PFR Task used by Wilks et al. (2021). The results from this stage can therefore be compared with those from Wilks et al. (2021) as a means of validating the current implementation of the PFR Method. We included Stage D (a 4 x 4 grid) to test whether it offered increased sensitivity for identifying potential for ratcheting. Stages A and B can be used to scaffold the learning of the focal Stages C and D, where required. This is likely to be necessary with non-human primates to develop their understanding of the task strategy prior to exposure to the more difficult stages (C and D) where it is harder to earn rewards when proficiency is low. This was not essential in the current study as verbal instruction can be provided to the children.

Each problem involves a demonstration trial followed by a single test trial. Figure 3.1 displays example trials from each of the four stages. Each individual selection is either rewarded (target found) or unrewarded (no target found).

Figure 3.1. PFR example trials from each stage (A-D) and cue type (visible cues = left panel; transient cues = right panel; see column 3). *The black and white striped circles are the targets, and black squares are non-targets. The first column of each panel displays the demonstration trial where selections are made by an animated blue hand; column 2 shows the demonstration trial output which is displayed until the continue button (green panel) is selected; column 3 shows the screen subjects made their choices from during the test trial, with either visible or transient cues; column 4 shows the final screen once all selections had been made, and displays example outcomes.*

The optimum strategy is to repeat rewarded demonstration trial selections and to avoid repeating those that were unrewarded by choosing alternative stimuli. As such, all pieces of information provided must be used strategically so that whether to copy or avoid can be deduced for each individual demonstration trial selection. In order to infer potential for ratcheting at the individual level (undertaken with non-human primates, Chapters 4 and 5), subjects must be exposed to multiple demonstrations of each possible performance score; this is done to generate enough power for analyses regarding outperformance of demonstrations. For the current chapter, this assessment was done at the group level (age bandings), so multiple exposures of each demonstration score per individual were not required. However, multiple data points were obtained from each individual in that each participant was exposed to a single trial for each possible performance score for both Stage C and D (e.g., at Stage C, each participant was exposed to four trials scoring *0*, *1*, *2* and *3*), and each participant completed this over both conditions (the consecutive transient and visible cues blocks, see Method).

If performing optimally, subjects had the potential to outperform the demonstrations they were exposed to on average (except when all targets were found in the demonstration trial). Figure 3.2 displays example demonstration and test trial outcomes for Stages C and D where the correct strategy either was, or was not, employed optimally.

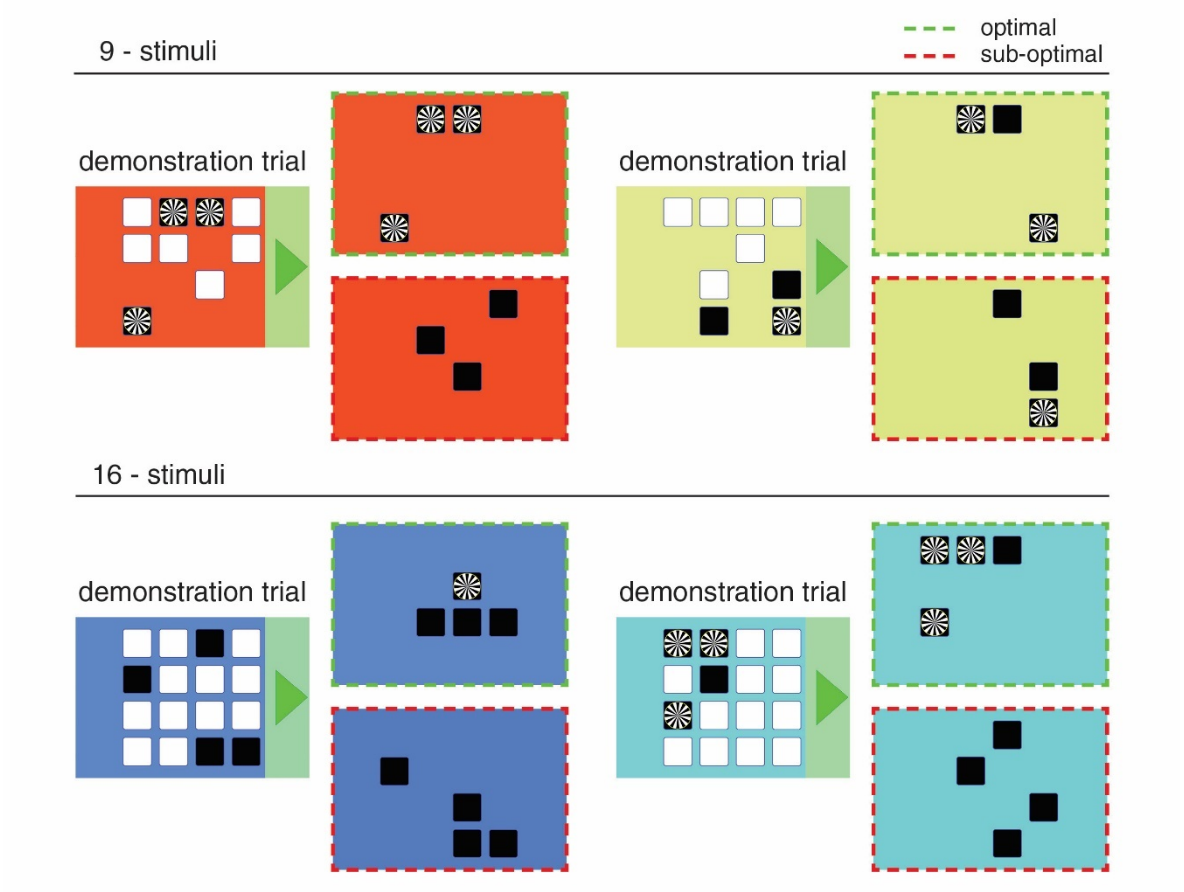


Figure 3.2. Example demonstration trial and test trial outcomes from Stages C (9-stimuli) and D (16-stimuli). *Test trials with a green dashed outline indicate optimal responding (repeating rewarded, avoiding unrewarded demonstrated selections) and the red dashed outline indicates sub-optimal responding (i.e., not repeating all rewarded or avoiding all unrewarded selections).*

Cue condition can also be manipulated to vary the role of memory, as in Wilks et al. (2021). In the visible cues condition, a mark indicating the outcome (rewarded or unrewarded) of each demonstration trial selection made remained on display while the subject made their choices; no such markers were provided in the transient cues condition during the test trial (Figure 3.1). This manipulation models the real-world variation in the availability of cues indicating the outcome of the activity of others (discussed in Chapter 1, section '*What Makes Human Culture So Unique?*'). Such cues may be of use to social learners.

To determine whether potential for ratcheting has been demonstrated, responses to the PFR Task are then used to assign a level to each age banding's (children, current chapter) and individual's (non-human primates, Chapter 4 and 5) capacity for cumulative culture according to a PFR Classification System (Table 3.2); also developed according to the logic of Caldwell et al. (2020). As previously noted, the aim of the current chapter is to replicate the pattern of results found by Wilks et al. (2021) who had also implemented a PFR Classification System based on the same logic. As such, we adopted a similar system to Wilks et al. (2021) to assign potential for ratcheting based on linear transmission involving a continuum of four levels. To be assigned to a particular level, all criteria associated with that level (outlined in Table 3.2) must be met. This is based on the aggregated performance on all trials performed by the group or the individual in question.

Table 3.2. PFR Classification System criteria based on a linear transmission scenario.

PFR Level	**Criteria (*Stage C, 9-stimuli; Stage D, 16-stimuli*)**
0	Chance level performance: • displays no benefit from the demonstration trial • performs at chance-level *(1/3; 1/4)* on average across all demonstration levels
1	Benefit but not associated with demonstration score: • performs above chance-level *(> 1/3; > 1/4)* overall • performance does not improve with increasing demonstration score; no significant difference in average performance following demonstrations that provided chance-level *(1/3; 1/4*), and higher-value *(2/3; 2/4, 3/4)* information
2	Benefit related to demonstration score, but not outperforming: • performs above chance *(> 1/3; > 1/4)* overall • higher scores as demonstration score increases • does not consistently outperform higher levels of demonstration *(2/3; 2/4, 3/4)*
3	Potential for Ratcheting: • higher scores following higher success demonstrations *(2/3; 2/4, 3/4)* compared to lower *(1/3; 1/4)* • on average, outperformance of both chance-level *(1/3; 1/4)* and the subsequent level of demonstration (2/3; 2/4)

Individuals that are essentially oblivious to the demonstration would perform at chance level and populate Level 0. This represents situations where the information has not been used at all. Level 1 involves performance that is above chance but that displays no measurable difference following lower and higher scoring demonstrations. Although this displays some sensitivity to the demonstration trial, this would be minimal, resulting in little retention of beneficial information and with no potential for increasing benefit over multiple transmission episodes. Level 2 classifies individuals or groups that are sensitive to the increasing value of the demonstration trials, but that fail to outperform the demonstration score. Although there is the potential for response patterns such as this to accumulate beneficial information under certain conditions, this could not occur under unbiased linear transmission (i.e., learning from a single, randomly selected individual from the previous generation).

Finally, to be classified as displaying PFR, Level 3 introduces an additional criterion involving outperformance of both chance and the following (consecutive) level of demonstration. This pattern of responding could theoretically support ratcheting in a linear transmission chain. Outperforming chance-level scores indicates the potential for improvement following observation of a task naïve participant's attempt. Continuing improvement is captured by the second criterion, which indicates increased success following receiving information from learners who had themselves benefited from exposure to chance level information. This type of performance would indicate that a ratchet effect could get off the ground.

In addition to assessing potential for ratcheting using the classification system, the results of the PFR Task can also be used to simulate the outcome of multi-generational transmission, as outlined in Chapter 1 (section '*The PFR Method*') and Caldwell et al. (2020). Briefly, the score for each 'generation' is produced by sampling from the pool of participant scores that were achieved in response to demonstrations whose value matches the score of the previous simulated generation (described in more detail in the Method below).

By running multiple simulations, the average score of each 'generation' following a simulated transmission event can then be calculated and visualised as a chain of learners. As such, the key alterations the PFR simulation makes compared to typical transmission chains is that the participants are 'exposed' to their own outputs, rather than requiring a unique individual at every generation. Thus, a multi-generational effect can be derived given individual level input.

Either the PFR Classification System can be used, or simulations can be conducted, however both of these approaches can be implemented to gain a more thorough understanding of potential for cumulative culture in a particular population or individual.

Predictions

Overall, we predicted that age and cue condition (visible vs transient) would affect performance, with greater success in older age groups, and in the visible cues condition. We further predicted success to increase with increasing demonstration score, and again that this would be more pronounced in older age groups and in the visible cues condition. Finally, we expected that both the top level met according to our PFR Classification System, and the top score attained when transmission chains were simulated would increase with age. This pattern of results would replicate the findings of Wilks et al. (2021) and would provide reassurance that the current implementation of the PFR method, like that used by Wilks et al. (2021), does allow for the identification of potential for ratcheting, whilst also picking up expected performance contrasts resulting from differences in cognitive load and test population. This would therefore provide an important validation of the current PFR Task for use with non-human primates.

Method

Participants

A total of 55 children were included in this study aged between two and seven years old (31 females; mean age = 56.69 months; SD = 18.72). Table 3.3 displays the number of children in each year-group banding; one 2-year-old took part in the visible condition only. A further seven children were tested but excluded because of missing or incorrect age information (N = 3) or exceeding our target age range (N = 5). The children were recruited at the University of Stirling's Summer Science Festival and a primary school in Scotland, and the majority of participants were British (N = 51).

Table 3.3. Sample size for each year banding.

Age	**2**	**3**	**4**	**5**	**6**	**7**
N	9	7	16	10	5	8

Ethical approval was granted by the General University Ethics Panel (GUEP) for the University of Stirling (reference: GUEP 629). Informed written consent was provided by the parent or guardian.

Materials

The children were presented with the training and task programme on a Microsoft Surface touch-sensitive capacitive tablet connected to a Microsoft wireless keyboard. The programme was written on PsychoPy 1.84.2 (Peirce et al., 2019) using Python2. Tokens, marbles, and stickers were used as positive reinforcement.

Procedure

Children were individually presented with the PFR Task in sessions that lasted approximately ten minutes. There were two consecutive experimental blocks: all problems in one block involved only visible cues and the other displayed only transient cues. Their order was counterbalanced. The experimenter gave verbal instruction before the commencement of each block (see the full script in Appendix 1).

Each block was preceded by two Stage B (Table 3.1) training problems of the same cue condition; problems involved a demonstration trial followed by a test trial (Figure 3.1). This was to familiarise the children with both the underlying task structure and the current cue condition. Children were then presented with Stage C and D test problems. The task was programmed to include one presentation of each possible demonstration score per stage, per block; these were ordered randomly. Therefore, each participant received nine problems per block involving four Stage C problems (scoring *0, 1, 2* and *3*) and five Stage D problems (scoring *0, 1, 2, 3*, and *4*), for a total of eighteen test problems overall.

The demonstration trial involved an animated blue hand that entered from the bottom-centre of the screen, and selected from an array of identical white, square-shaped stimuli (Figure 3.1). Table 3.1 indicates the number of stimuli (2-16), selections made during each trial (1-4), and the number of targets available (1-4) at each stage (A-D) of the PFR Task. Each selection made was either rewarded, whereby an arbitrary target (black and white striped circle – described as 'wagon wheels' to the children) replaced the chosen stimulus, and an arbitrary sound was played, or unrewarded, where a black square replaced the chosen stimulus, and a different arbitrary sound was played. Children were explicitly instructed to find as many targets as possible. Once all selections had been made, the animated hand exited at the bottom-centre of the screen and a 'continue' button immediately appeared on the right side of the screen (green arrow/panel). The outcome of all demonstration trial selections was displayed until the button had been selected. The continue button was included to minimise the issue that timed displays are subject to attentional fluctuations, effectively removing the potential for participants to 'miss' the outcome of the demonstration due to lack of attention at a critical time.

Activation of the continue button initiated the test trial where the participants were presented with the same array of stimuli to choose from. The locations of the stimuli and the targets remained constant within each individual problem. The successful strategy was to repeat rewarded stimuli selections and to avoid reselecting unrewarded stimuli by choosing an alternative stimulus (Figure 3.2). Once all selections had been made, the test trial outcome (whether each selection was rewarded or unrewarded) was displayed for three seconds. The screen then turned black until the next problem was initiated.

On Stage D the stimuli were arranged in a 4 x 4 grid (Figure 3.1). However, on Stage C and practise Stage B (and Stage A used in Chapters 4 and 5), the stimuli could be located at any position on the same 4 x 4 grid arrangement; these locations were randomly selected between problems. This meant that, for example, on Stage C, the locations of each of the 9 stimuli were randomly assigned from the sixteen possible options on the underlying 4 x 4 grid (unused locations were not visible; see Figure 3.1 A-C). This was done to encourage use of all areas of the screen, with the aim to minimise development of potential location biases with non-human primates. This randomisation of positioning is a departure from the design of the study by Wilks et al. (2021). This, together with the other key methodological differences (electronic presentation, task framing [finding treasure vs. finding wagon wheels in the current study]), formed the basis of our motivation to replicate Wilks et al. (2021) to validate the current task.

The background colour also randomly varied between problems. This was to reinforce the conception of separate problems that each had their own reward structure and also to highlight the correspondence between the demonstration and test trial *within* each individual problem. On Stages A-C, this was also reinforced by the variation in the location of the stimuli between problems.

At the end of the study all children received a sticker, and children recruited at the festival were given a token that could be exchanged for a range of leisure activities.

Analysis

Four general linear mixed models (GLMM) were constructed to test our predictions relating to PFR Task performance using the *glmer* function of the *lme4* package (Bates et al., 2014) on R (R Core Team, 2020). Models one and three analysed the results of Stage C (9-stimuli), and models two and four analysed Stage D (16-stimuli) data.

GLMM one and two included *repetition* as a binary outcome variable, where an individual selection of any stimulus chosen during the demonstration was recorded as ‘repetition’ [1] and selection of any other stimulus was recorded as a ‘shift’ [0] response. Age (in days, centred and scaled), information type (rewarded vs unrewarded) and cue type (visible vs transient), were included as fixed effects, as well as their interactions.

Models three and four used *task score* as a binary outcome variable, recording any selection that discovered a target as successful [1] and any selection that did not as unsuccessful [0]. This was different to the above outcome measure as it incorporated the element of chance involved in discovering the location of targets not revealed during the demonstration trial. Age (in days, centred and scaled), cue type, and demonstration score (0-3 on Stage C; 0-4 on Stage D) and their interactions were included as fixed effects.

When included in any model, information type, cue type, and demonstration score were sum coded so that their levels were -1 and 1. When included as fixed effects, information type, cue type, and demonstration score were also included as random slopes. Each model included participant ID as a random intercept. As described in Chapter 2, ‘*Analysis*’ section, a maximal effects structure was used. For the task score GLMMs, cue type was dropped as a random slope on the Stage C model, and demonstration score was dropped as a random slope on the Stage D model due to non-convergence.

In the event of significant interactions, estimated marginal means were computed using the *emmeans* package (Lenth et al., 2019) on R Studio as post-hoc tests to ascertain the direction of the effects. These results are reported on the log odds ratio scale. When age and demonstration score were included in any post-hoc analyses, comparisons between older vs younger children, and high vs low scoring demonstrations were calculated using the upper and lower quantiles, respectively.

For assessing potential for ratcheting *only*, and for the purpose of comparing our results to Wilks et al. (2021) (who grouped children aged 3-7 years into four bandings of 3-year-olds, 4-year-olds, 5-year-olds, and 6-year-olds, with the oldest group also including one 7-year-old) , we also grouped children into four age brackets. The age brackets used here were 2-to-3-year-olds (age range = 25 - 45 months, *N* = 16), 4-year-olds (age range = 48 - 58 months, *N* = 16), 5-year-olds (age range = 61-69 months, *N* = 10), and 6- to 7-year-olds (age range = 70 - 90 months, *N* = 13).

To assess whether each age bracket met the criteria for Levels 1 and 3 of our PFR Classification System (Table 3.2) for each cue condition, we calculated the mean task score overall, and for each demonstration score (1 and 2 for Stage C, and 1, 2 and 3 for Stage D) respectively. One-sample t-tests were conducted to assess whether PFR Classification Level 1 was met with significance. Pearson's correlations were performed to assess whether there was a significant, positive linear relationship between demonstration score (1-3 for Stage C, and 1-4 for Stage D) and task score, and thus whether Level 2 was met.

Inferential analyses were not performed to assess the requirements for Level 3. The low participant numbers in each grouping (as here analyses were performed on responses separated by demonstration score), compared to the larger numbers obtained by Wilks et al. (2021) would have necessarily generated varying *p*-values between the two studies, whereas means can be directly compared. Moreover, numerical outperformance of higher scoring demonstrations would be unlikely by chance, thus a numerical improvement is informative for Level 3.

However, significance was required to achieve Levels 1 and 2. Firstly, there was enough power for these statistical tests as they involved a greater number of responses per test. In addition, this was necessary at these levels as overall scores that were numerically above chance and correlations that were simply positive may be expected on this task, as these criteria may be easily met even if responding at random.

The PFR Task responses of our age brackets were also used to simulate linear transmission chains for each array size and cue condition, for a total of sixteen simulations (more detail below).

Results

Repeats GLMMs

Stage C: 9-stimuli

The maximal model for Stage C was significantly better than a null equivalent (χ^2 (7) = 127.09, *p* < .001) (Figure 3.3). Children were found to repeat rewarded selections more than unrewarded (*b* = 3.10, SE = 0.31, z = 9.86, *p* < .001) indicating an overall grasp of the task contingencies. There was no significant main effect of cue type (visible vs transient) (*p* = .62).

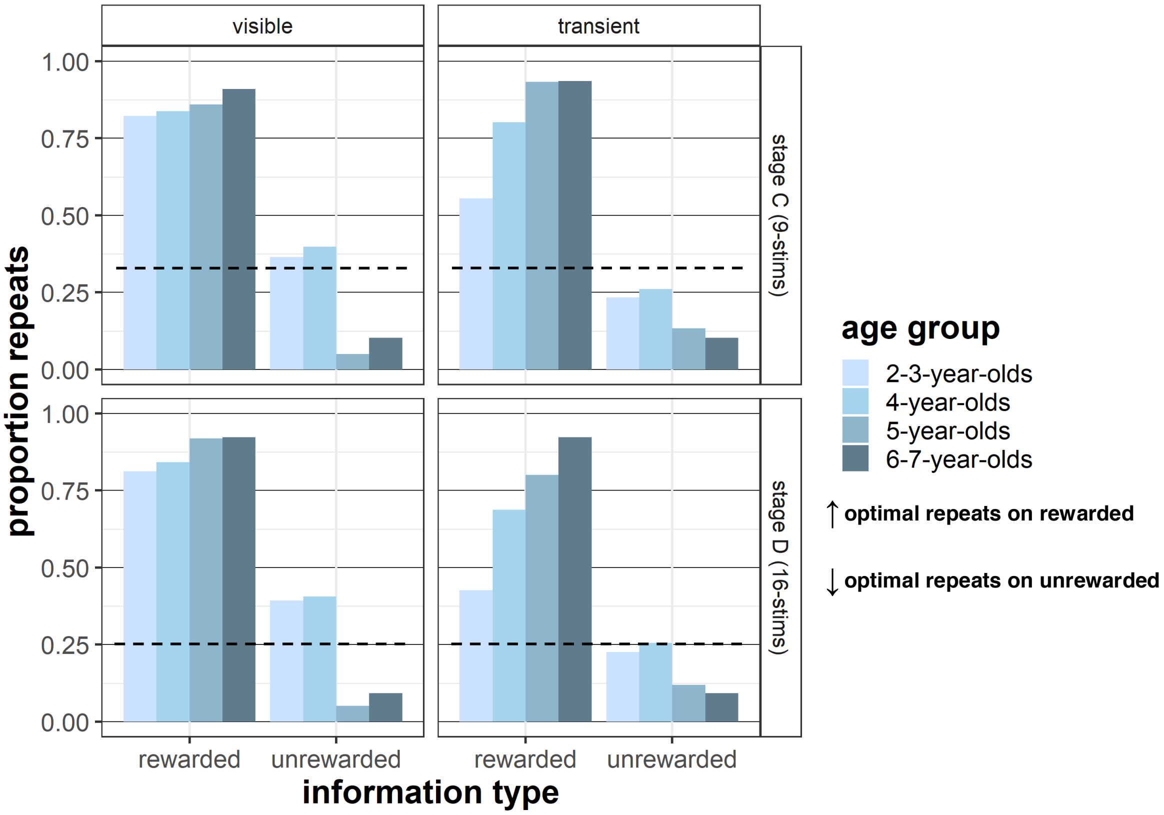


Figure 3.3. Average repetition by information type and age, faceted by cue type and stage. *Age is grouped for illustration purposes only. Dashed lines indicate chance of repetition based on random responding (33% on Stage C; 25% on Stage D).*

Although there was no significant main effect of age (*p* = .60), there was an interaction between age and information type (rewarded vs unrewarded) (*b* = 1.31, SE = 0.26, z = 5.12, *p* < .001). An *emmeans* post-hoc analysis clarified that, compared with the younger children, the older children repeated more when information was rewarded (*b* = 1.67, SE = 0.60, z = 2.78, *p* = .006) and less when information was unrewarded (*b* = 2.18, SE = 0.63, z = 3.47, *p* < .001) (Figure 3.4). This pattern indicates a better grasp of the optimal task strategy with age.

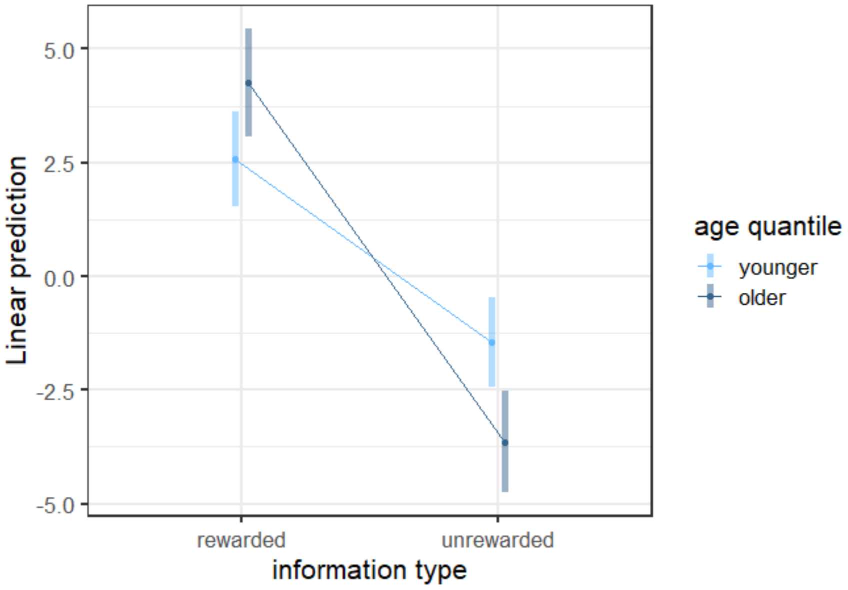


Figure 3.4. Interaction of age and information type on repetition on Stage C.

There was also an interaction between age and cue type (*b* = 0.61, SE = 0.27, z = 2.28, *p* = .02). A post-hoc analysis found a significant difference between older and younger children when cues were transient (*b* = 0.65, SE = 0.28, z = 2.27, *p* = .02) but not when cues were visible (*b* = 1.15, SE = 0.84, z = 1.37, *p* = .17) (Figure 3.5). This suggests that transient cues differentially affected repetition dependent on age (with older children repeating more than younger), but visible cues did not.

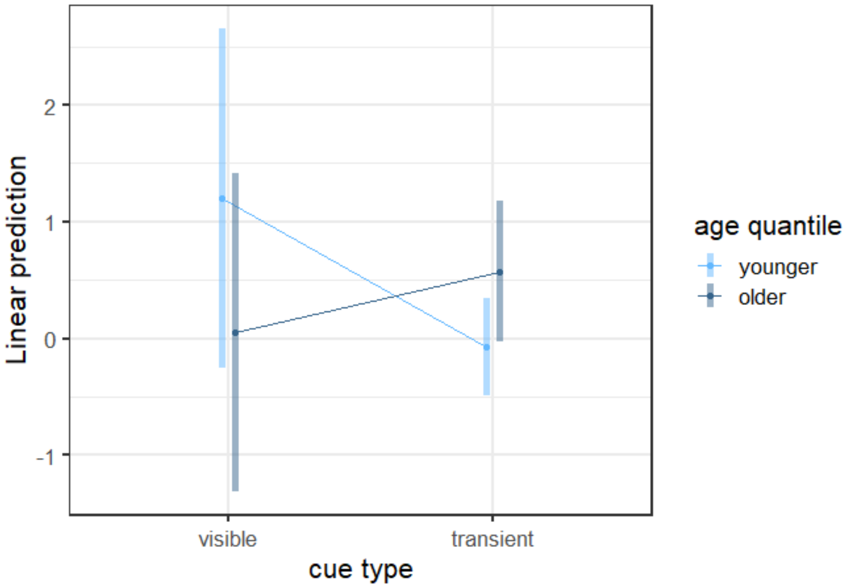


Figure 3.5. Interaction between age and cue type on repetition at Stage C.

Information type and cue type also interacted to affect repetition (*b* = 0.95, SE = 0.21, z = 4.43, *p* < .001) (Figure 3.3). An *emmeans* post-hoc test found that the effect of information type was slightly more pronounced when cues were visible (*b* = 8.09, SE = 0.94, z = 8.64, *p* < .001) compared to transient (*b* = 4.30, SE = 0.53, z = 8.14, *p* < .001) (Figure 3.6), suggesting better implementation of the task strategy when cues remained visible.

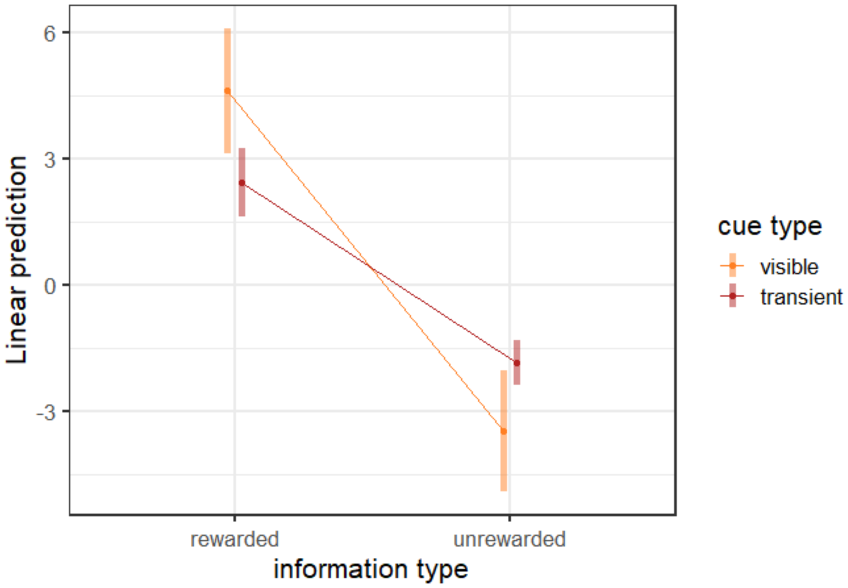


Figure 3.6. Interaction of information type and cue type on repetition on Stage C.

The 3-way interaction between age, information type and cue type was not significant (*p* = .26).

Stage D: 16-stimuli

The GLMM constructed for the Stage D data was also more predictive than the null (χ^2 (7) = 173.24, *p* < .001) (Figure 3.3). As above, there was an overall increase in repetition when demonstration trial selections had been rewarded, compared with unrewarded (*b* = 2.39, SE = 0.19, z = 12.61, *p* < .001) in line with the successful strategy. As on 9-stimuli arrays (Stage C), there was no significant main effect of cue type (*p* = .09) or age (*p* = .87) on repetition.

Again, age interacted with information type (*b* = 1.21, SE = 0.17, z = 7.01, *p* < .001) to affect repetition. Post-hoc analysis again revealed that older children repeated significantly more in response to rewarded information (*b* = 1.73, SE = 0.42, z = 4.08, *p* < .001) and less when the information was unrewarded (*b* = 1.83, SE = 0.40, z = 4.57, *p* < .001), relative to younger children.

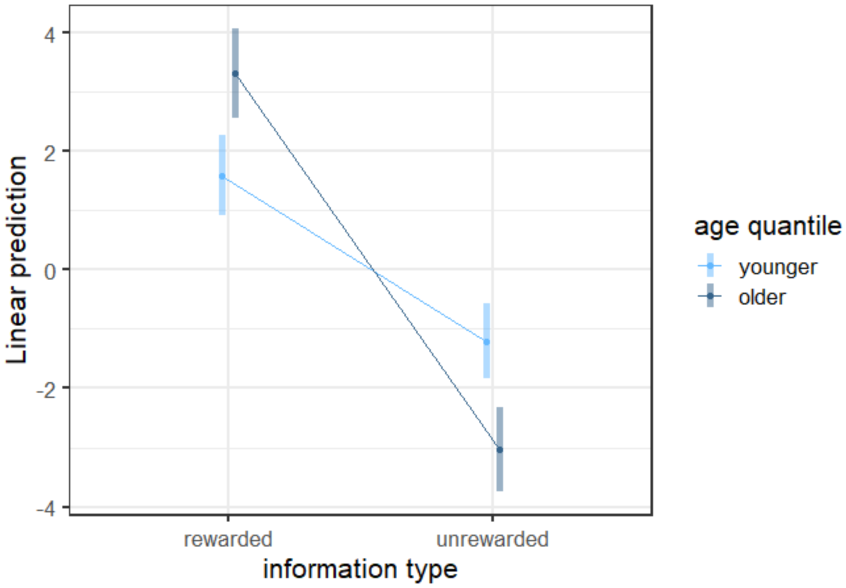


Figure 3.7. Interaction of age and information type on repetition on Stage D.

A significant interaction was also found between age and cue type (*b* = 0.46, SE = 0.21, z = 2.23, p .03). As on Stage C, post-hoc testing found that older children repeated more than younger children when cues were transient (*b* = 0.62, SE = 0.20, z = 3.16, *p* = .002) however no significant difference was found when cues were visible (*b* = 0.72, SE = 0.60, z = 1.22, *p* = .22) (Figure 3.8).

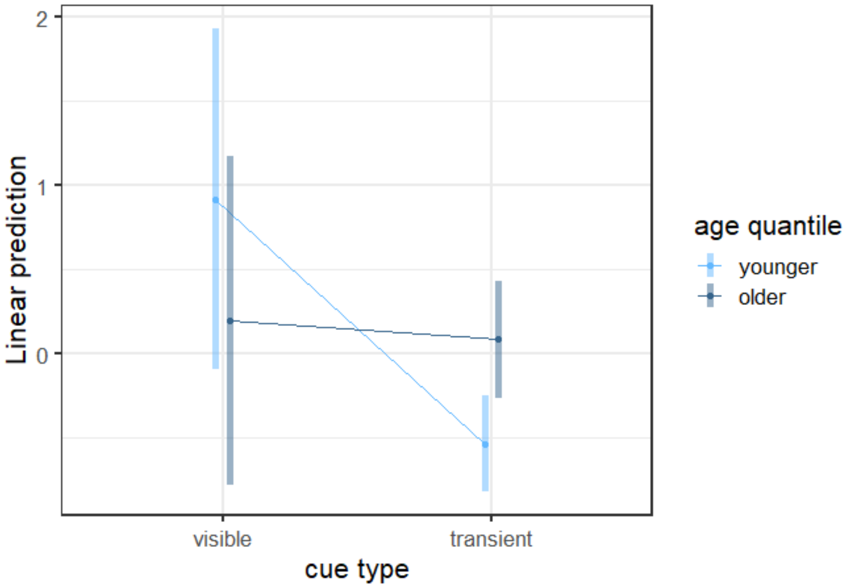


Figure 3.8. Interaction of age and cue type on repetition at Stage D.

There was also an interaction between information type and cue type (*b* = 0.87, SE = 0.14, z = 6.08, *p* < .001). Post-hoc results indicated that the effect of information type was more pronounced when cues were visible (*b* = 6.51, SE = 0.59, z = 10.96, *p* < .001) compared to transient cues (*b* = 3.04, SE = 0.31, z = 9.77, *p* < .001) (Figure 3.9), also corresponding to Stage C.

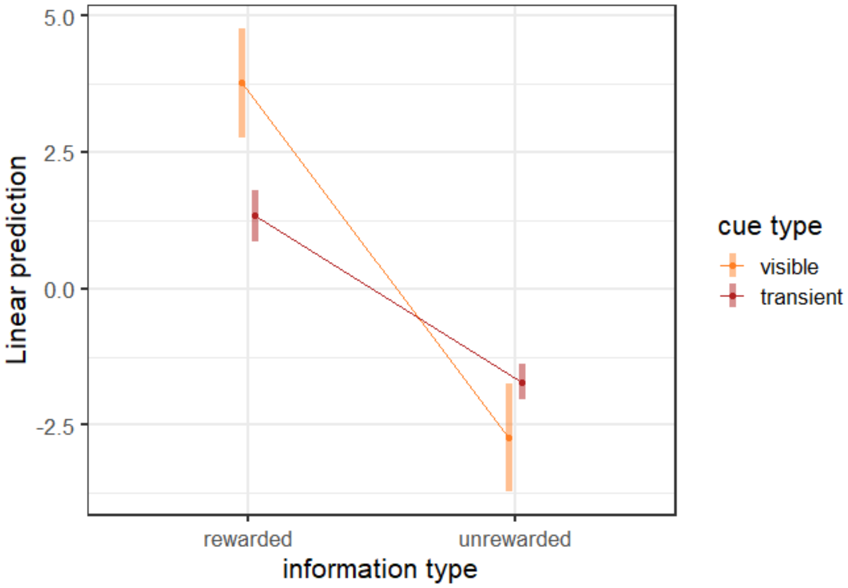


Figure 3.9. Interaction of information type and cue type on repetition on Stage D.

Finally, unlike on the Stage C data, there was a significant 3-way interaction of age, information type and cue type (*b* = 0.25, SE = 0.12, z = 2.10, *p* = .04). Post hoc analysis revealed a significant effect of cue-type when information was rewarded in both age quantiles (upper: *b* = 2.17, SE = 0.67, z = 3.26, *p* = .001; lower: *b* = 2.77, SE = 0.57, z = 4.88, *p* < .001) and when information was unrewarded in older children (*b* = 1.96, SE = 0.64, z = 3.07, *p* = .002), but there was no variation by cue type in the younger children when information was unrewarded (*b* = 0.13, SE = 0.54, z = 0.24, *p* = .81). This suggests that visible cues were linked to an improvement in the use of rewarded information overall, and in the older children when information was unrewarded. However, younger children performed similarly across the cue conditions in response to unrewarded information.

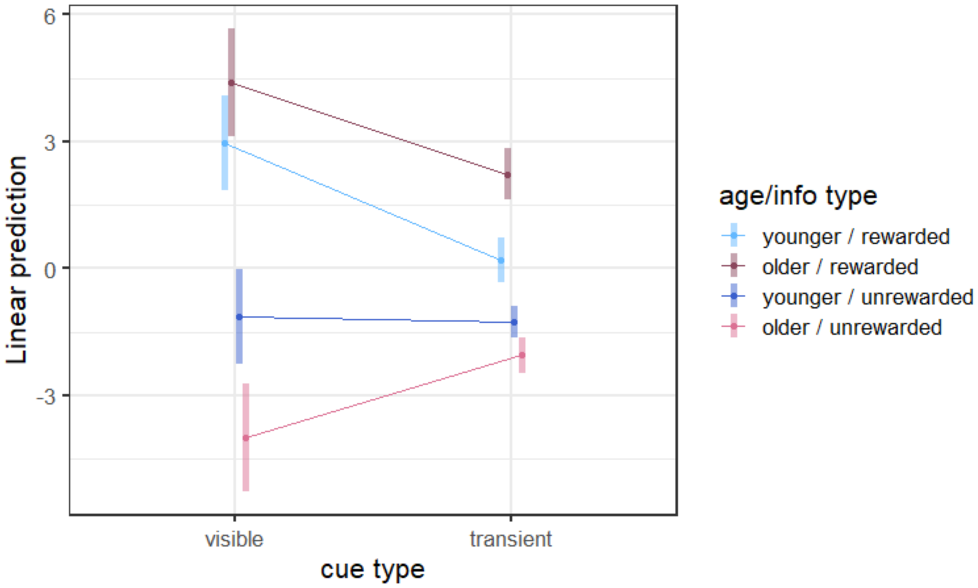


Figure 3.10. Interaction between age, information type and cue type affecting repetition on Stage D.

Task Score GLMMs

Stage C: 9-stimuli

The GLMM built for task score (number of targets found) on the Stage C data was more predictive than the null (χ^2 (7) = 79.79, *p* < .001) (Figure 3.11). Increasing demonstration score was associated with higher task scores overall (*b* = 0.67, SE = 0.08, z = 8.06, *p* < .001). There was no significant main effect of cue type (*p* = .58), and the main effect of age approached significance (*p* = .052).

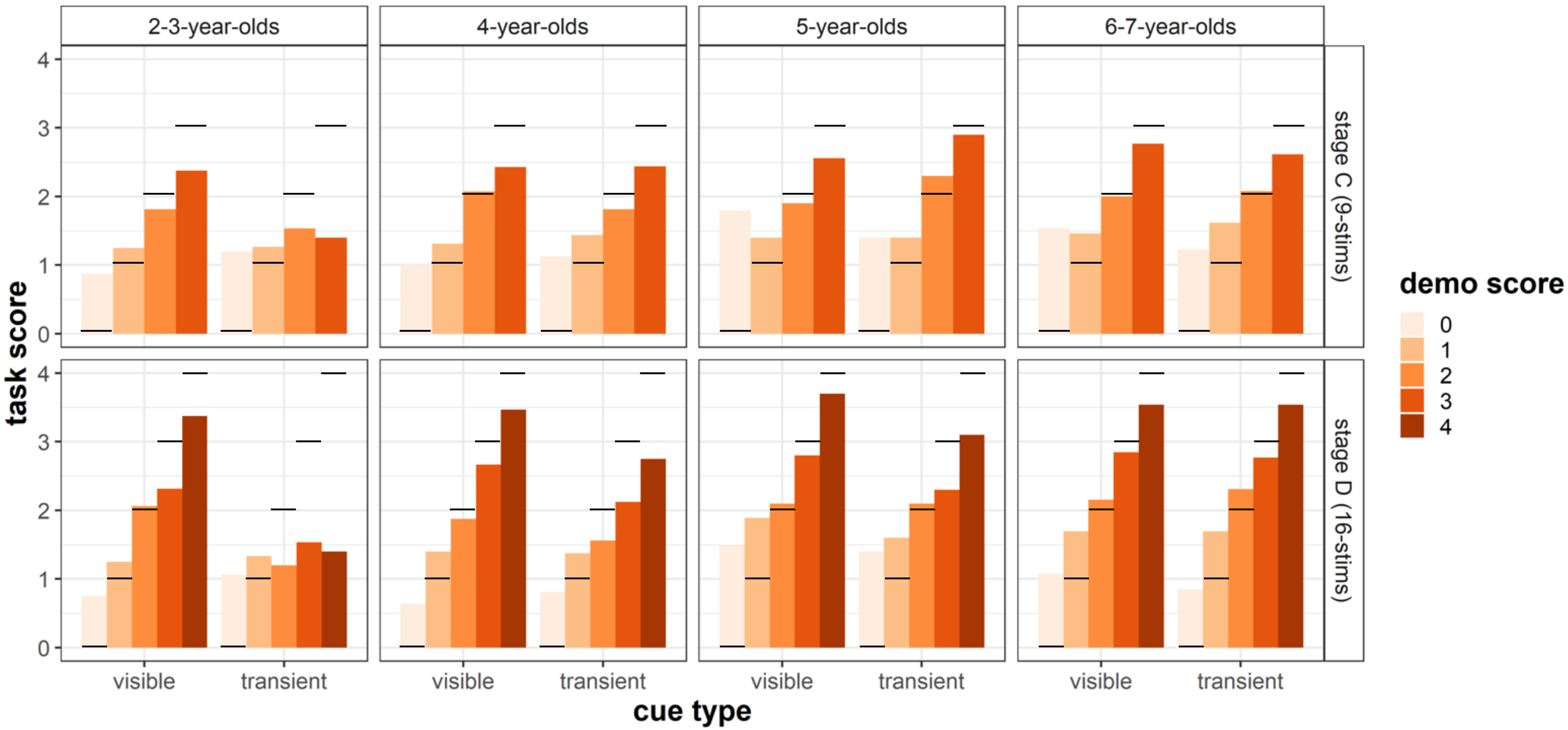


Figure 3.11. ***Mean task score (number of targets found) by cue type, demonstration score and age.*** *Age is grouped for illustration purposes only. Horizontal black lines highlight the associated demonstration score level.*

The interaction between age and cue type was a significant predictor of task score (*b* = 0.20, SE = 0.10, z = 1.97, *p* = .048). A post-hoc analysis found a non-significant trend toward lower task scores in the younger children when cues were transient compared to visible (lower quantile: *b* = 0.24, SE = 0.16, z = 1.52, *p* = .13), whereas the trend (also non-significant) was in the opposite direction for the older children (upper quantile: *b* = 0.05, SE = 0.15, z = 0.32, *p* = .75) (Figure 3.12).

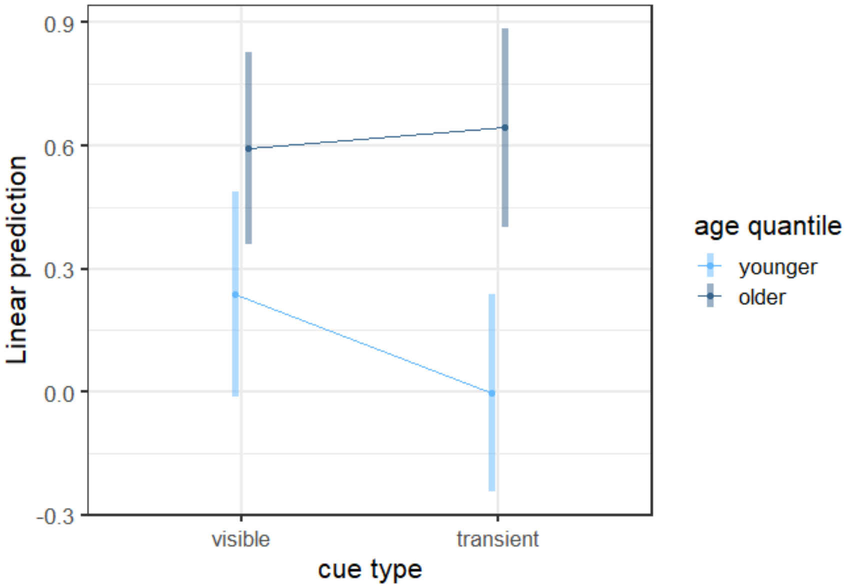


Figure 3.12. Interaction of age and cue type on task score on Stage C.

A significant 3-way interaction between age, demonstration score and cue type was also found (*b* = 0.20, SE = 0.06, z = 3.38, *p* < .001). A post-hoc test revealed a significant difference between the cue conditions in younger children when demonstration value was high (upper quantile: *b* = 0.47, SE = 0.18, z = 2.62, *p* = .009) but not when demonstrations score was low (lower quantile: *b* = 0.02, SE = 0.17, z = 0.09, *p* = .93) (Figure 3.13). No significant differences were found between cue types in the older children when demonstrations scored high (upper quantile: *b* = 0.12, SE = 0.18, z = 0.64, *p* = .52) or low (lower quantile: *b* = 0.02, SE = 0.16, z = 0.11, *p* = .92).

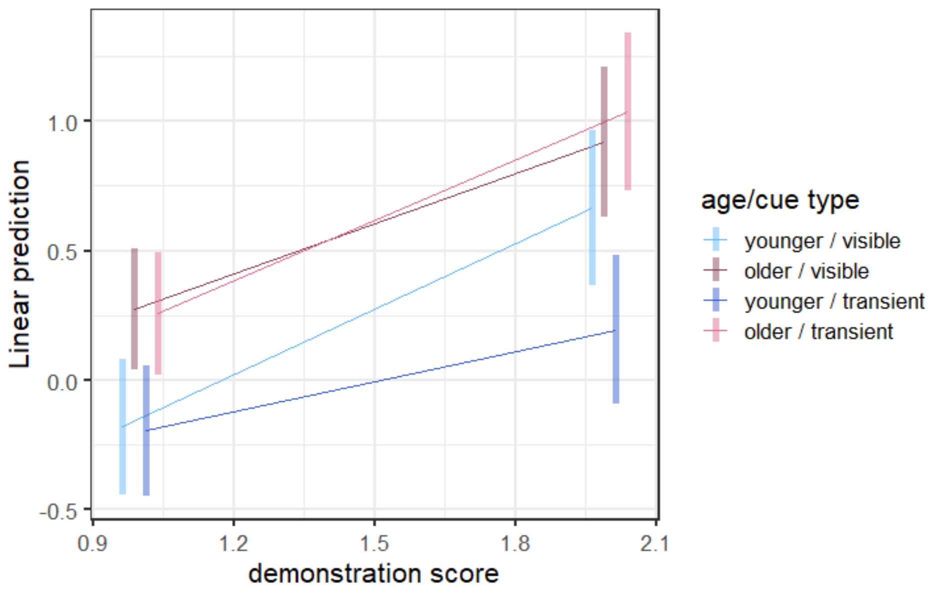


Figure 3.13. The three-way interaction of age, demonstration score and cue type on task score at Stage C.

The interactions between age and demonstration score (*p* = .43), and cue type and demonstration score (*p* = .29) were not significantly associated with task score.

Stage D: 16-stimuli

The task score GLMM on the Stage D dataset was also a better fit of the data than the null equivalent (χ^2 (7) = 382.01, *p* < .001) (Figure 3.11). As above, higher demonstration scores were associated with finding more targets (*b* = 0.60, SE = 0.04, z = 16.10, *p* < .001). Again, age (*p* = .19) and cue type (*p* = .20) were not significant predictors.

Unlike Stage C, the interaction between age and demonstration score was found to significantly affect task score on Stage D (*b* = 0.09, SE = 0.04, z = 2.49, *p* = .01). Further post-hoc analysis suggested that the effect of demonstration score was very slightly stronger in the older children (upper quantile: *b* = 0.66, SE = 0.45, z = 14.67, *p* < .001) compared to younger (lower quantile: *b* = 0.52, SE = 0.05, z = 10.96, *p* <.001) (Figure 3.14).

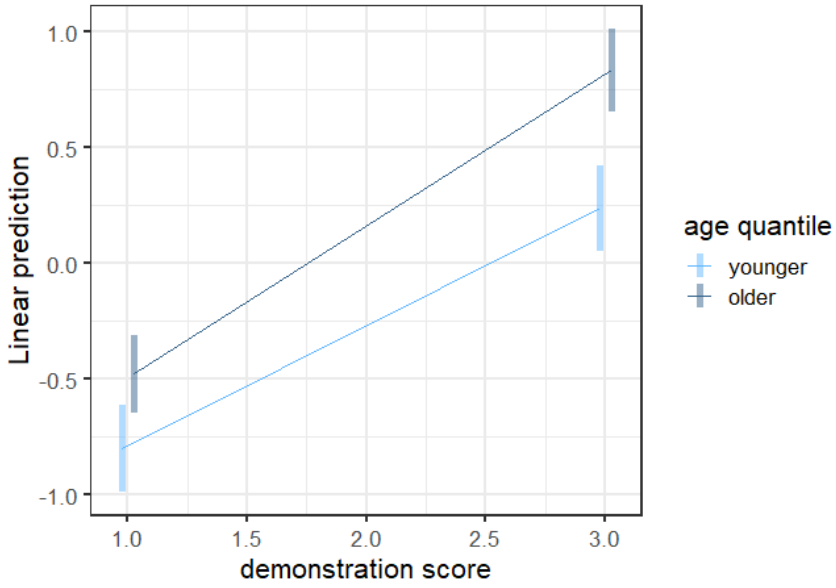


Figure 3.14. Interaction between age and demonstration score on Stage D.

Also in contrast to the Stage C model, there was a significant interaction between cue type and demonstration score (*b* = 0.15, SE = 0.04, z = 4.04, *p* < .001). Post-hoc results suggested that the effect of demonstration score was slightly stronger when cues were visible (*b* = 0.75, SE = 0.06, z = 13.50) than transient (*b* = 0.45, SE = 0.05, z = 9.05, *p* < .001) (Figure 3.15).

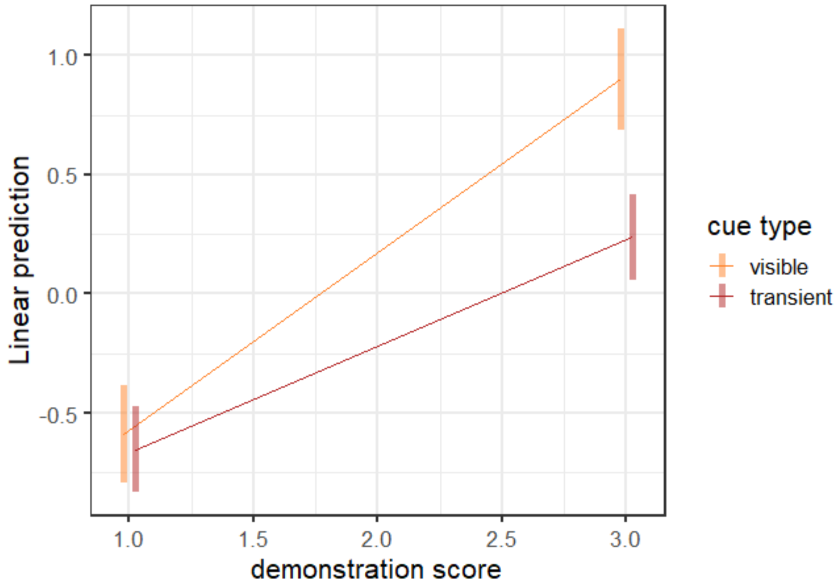


Figure 3.15. Interaction between cue type and demonstration score on task score on Stage D.

As above, the three-way interaction between age, demonstration score and cue type was also significant (*b* = 0.15, SE = 0.04, z = 3.90, *p* < .001). Post-hoc tests revealed a similar pattern to Stage C, as a significant effect of cue type was found only in the younger children when demonstrations scored highly (*b* = 0.60, SE = 0.14, z = 4.23, *p* < .001), all other combinations were not significant (young-low *b* = 0.07, SE = 0.17, z = 0.38, *p* = .71; old-low *b* = 0.07, SE = 0.15, z = 0.43, *p* = .67; old-high *b* = 0.18, SE = 0.13, z = 1.33, *p* = .18) (Figure 3.16).

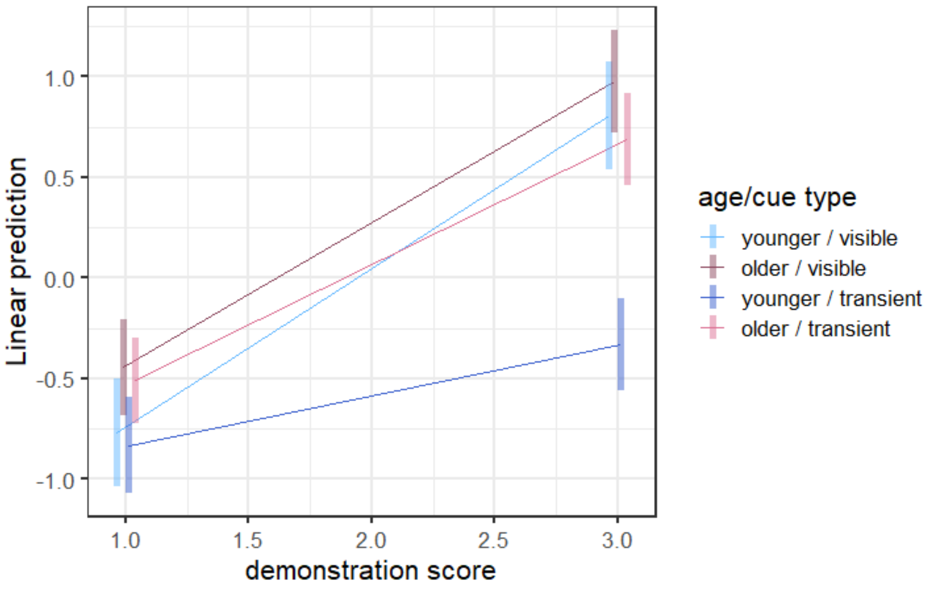


Figure 3.16. The interaction between age, cue type and demonstration score affecting task score at Stage D.

The interaction term between age and cue type (*p* = .28) was not found to be a significant predictor of task score at Stage D.

Potential for Ratcheting

As outlined in the introduction, we also utilised each age bracket's PFR Task responses to classify potential for ratcheting at a level between 0-3 (Table 3.2), and to simulate linear transmission chains for each array size and cue condition.

Essentially, we investigated whether i) overall task score (number of targets found) was higher than chance (Level 1), ii) task score was positively related to demonstrations score (Level 2) and iii) task scores were higher than the value of the associated demonstration in response to (at least) demonstrations scoring 1 and 2 (Level 3). The mean task score for each stage, cue type and level of demonstration for each age bracket can be found in Appendix 2 (also see Figure 3.11) as well as the t-test and correlation results.

Figure 3.17 displays the top PFR level met by each age bracket according to our classification system. Higher classifications appear to have been attained on visible cues by the youngest age bracket in comparison to transient cues. Little variation between the cue conditions is apparent in the older age brackets; however, an exception is on Stage C-visible cues, where slightly lower potential was displayed.

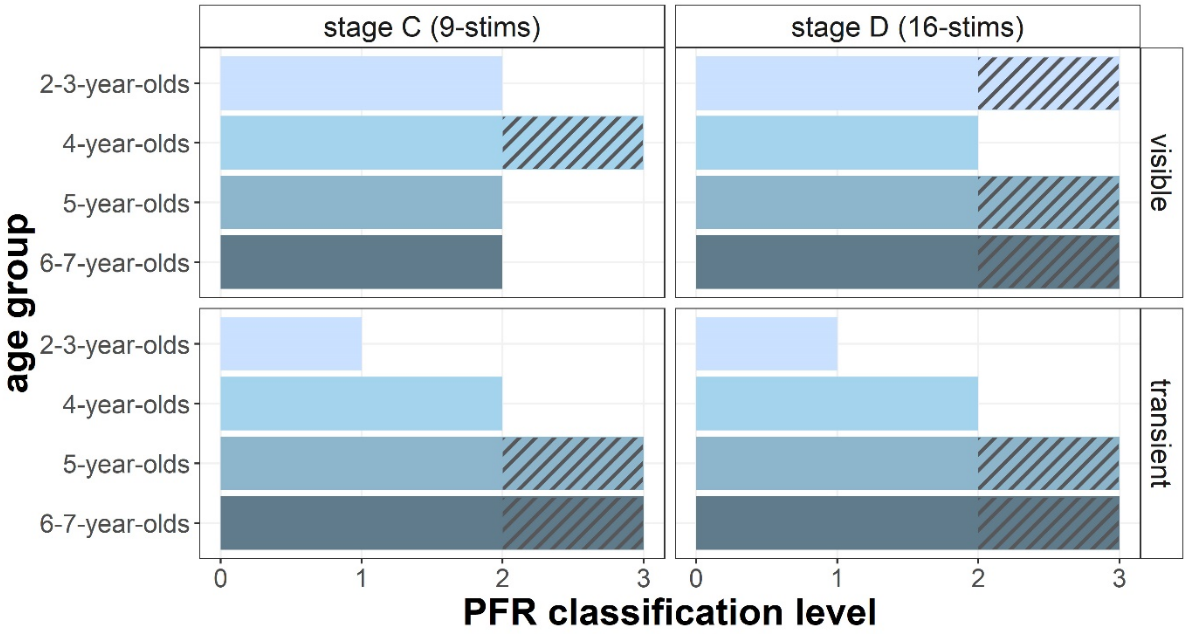


Figure 3.17. PFR classification level for each age grouping, cue condition and stage. Shaded sections indicate that Level 3 was numerically met only (without significance).

The youngest age category, 2- to 3-year-olds, reached Levels 2 and 3 on visible cues but were far from meeting even the Level 2 criteria on transient cues (Appendix 2, see also Figure 3.11). In contrast, in the conditions where Level 3 was not met by the 5- and 6- to 7-year-olds (Stage C-visible) average scores were relatively close to the necessary criteria for this level (Appendix 2), further demonstrating the lower variation between cue types in these age bandings. This will be considered further in the Discussion.

The lower classifications reached by the youngest age category on transient cues (compared to both the visible condition, and the oldest age brackets) are congruent with the greater spread of age effects found on transient in comparison to visible cues in the task score post-hoc analyses above (section ‘*Task Score GLMMs*’). Thus, transient cues appear to have severely inhibited potential for ratcheting in younger children.

On transient cues overall, the highest levels met did not appear to vary between the two stages. However, on visible cues, although there were some discrepancies, there was no clear pattern regarding the top level met between Stages C and D across our age brackets, suggesting that increasing complexity in this manner does not produce a systematic difference in classification levels.

Next, we simulated transmission chains to further examine potential for ratcheting abilities across the age categories. First, we pooled task scores (the number of targets found) according to the associated demonstration score, i.e., we grouped responses according to the value of the demonstration that had preceded them (for example, all responses given when demonstrations had scored *1* were grouped together, and responses to demonstrations that had scored *2* were grouped together, etc.). ‘Generation 0’ reflected performance in response to a blank grid displaying no cues, i.e., selecting at random. Individual chains were created by selecting from the pool of responses given following a demonstration of the same value as the previous generation’s score (Table 3.4). One thousand simulations were conducted for each cue condition, and each involved fifteen generations. We also created an ‘optimal’ chain, which indicates performance if the correct strategy was consistently implemented (i.e., rewarded always repeated, and unrewarded never repeated).

Table 3.4. Outline of chain construction. *Key:* G_x *= generation number; DS = demonstration score; s = task score.*

		example
G_0	randomly generated score	$G_0s = 1$
↓		↓
G_1	randomly select from **pool** of data where DS = G_0s	$G_1s = 2$
↓		↓
G_2	randomly select from **pool** of data where DS = G_1s	$G_2s = 2$
…		…
↓		↓
G_{10}	randomly select from **pool** of data where DS = G_9s	$G_{10}s = 3$

The average score of each position in the chain was then taken (Figure 3.18). The extent of accumulation that would be possible in a population, based on linear transmission, was estimated by inspecting where performance plateaued. Overall, older age groups performed better, compared with the younger groups, when responses were chained in the manner described. Visible cues were associated with higher top scores across both stages, compared with the transient cues, with the exception of the 5-year-olds on Stage C, where performance was substantially higher in the transient cues condition. Stage D appeared to pick up greater variation between the cue types, with larger differences in top scores between visible and transient cues for the 4- and 5-year-olds.

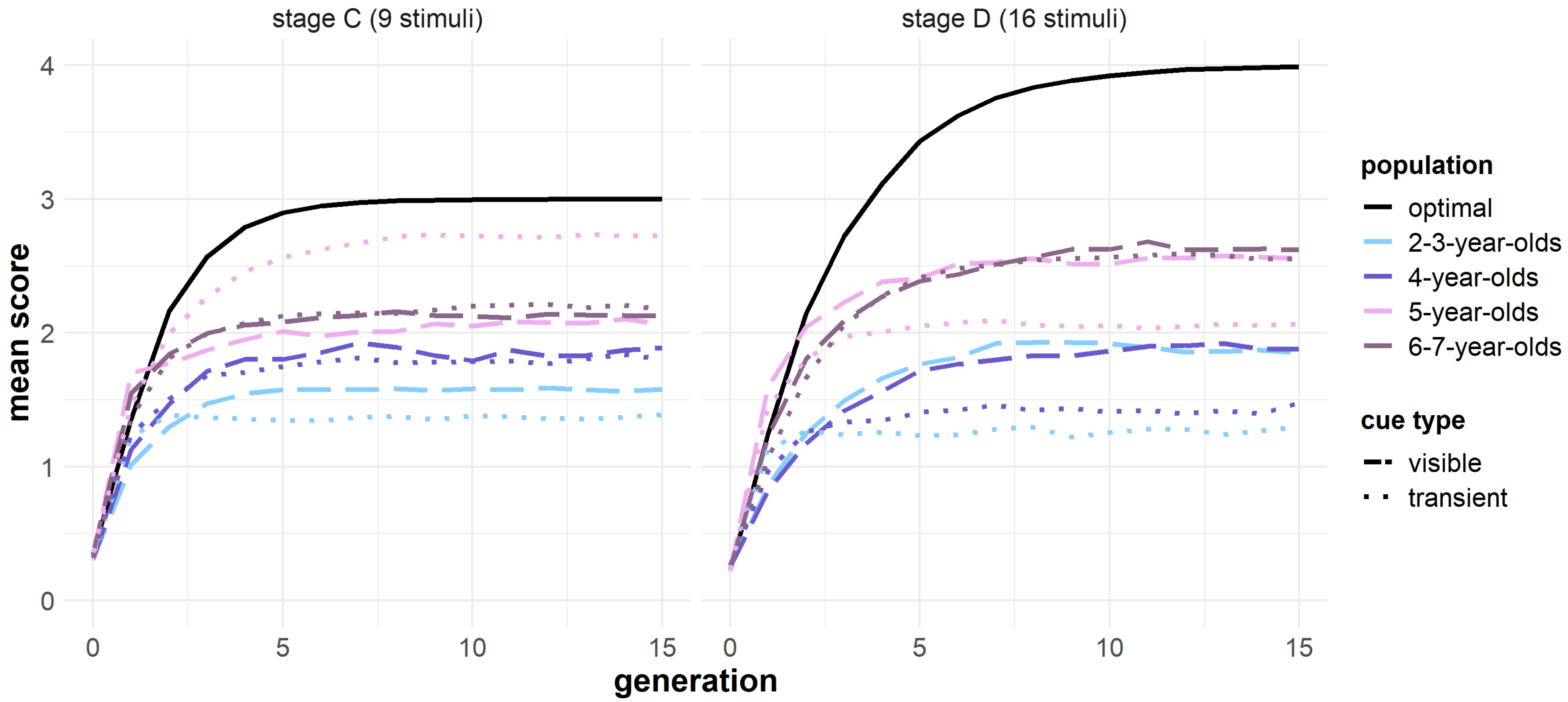


__Figure 3.18. Simulated transmission chain performance for each age grouping.__ Simulations from all cue conditions and stages (left = Stage C, right = Stage D) are displayed. The solid black line represents optimal performance i.e., fully consistent application of correct strategy. Chancel level for Stage C was 1/3 and for Stage D was 1/4.

With the possible exception of the 5-year-olds on Stage C-transient cues, all age brackets were far from reaching optimal scores regardless of age or cue condition. This sub-optimal performance pattern was more pronounced on Stage D and suggests low precision application of the optimal strategy. Figure 3.19 displays repetition across the range of demonstration values, highlighting the largely sub-optimal strategy use with no clear influence of demonstration score.

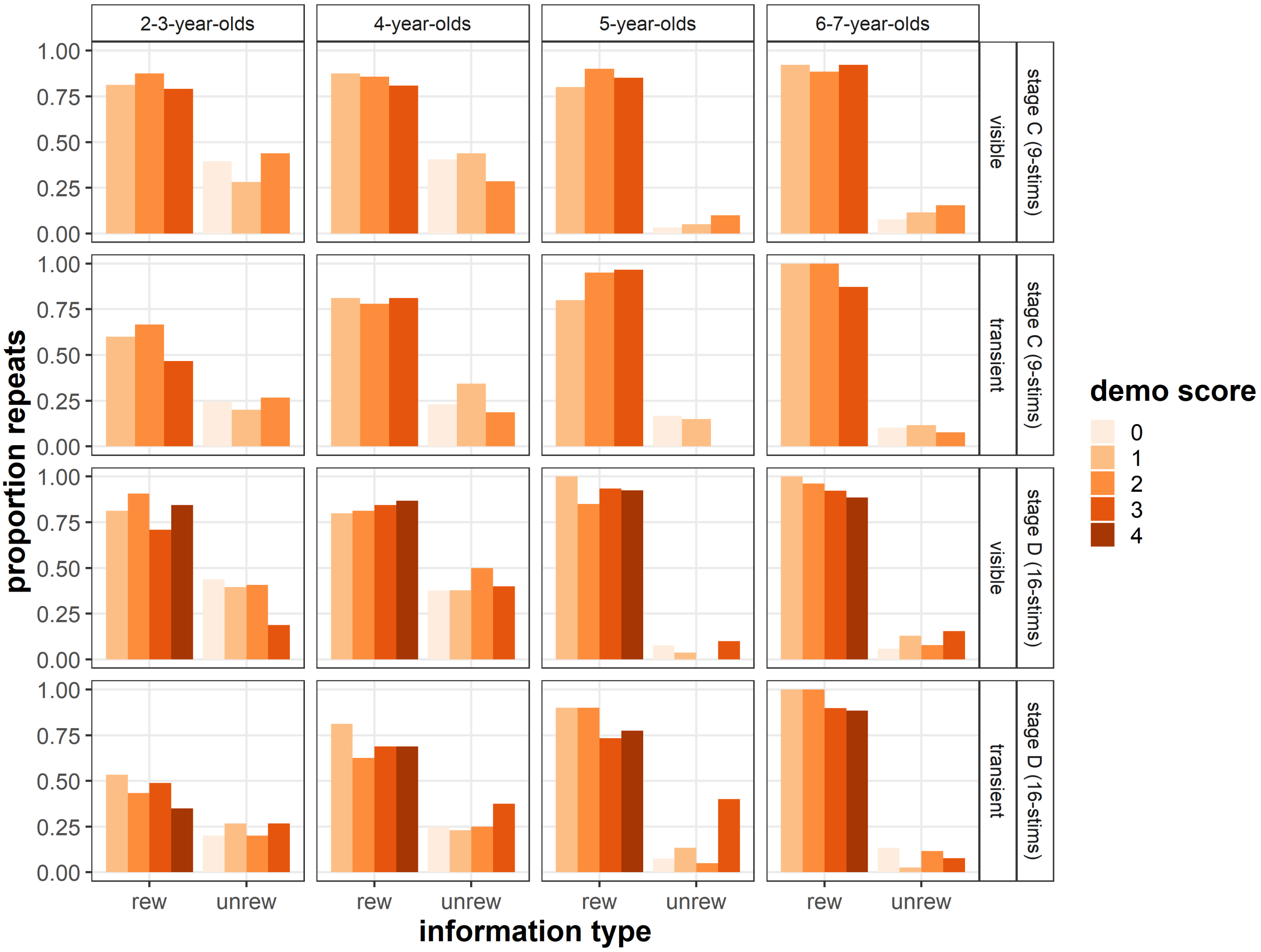


Figure 3.19. Proportion of repetition by information type (rewarded/unrewarded) coloured by demonstration score, and faceted by age banding, cue type and stage.

Discussion

In line with predictions, older children were found to employ the successful strategy (repeat rewarded, avoid unrewarded selections) more proficiently than younger children on both Stage C (9-stimuli) and D (16-stimuli). Younger children appeared to differentiate less well between information types (rewarded and unrewarded) when cues were transient compared to visible, on both stages. This was likely driven by the youngest year banding (2- to 3-year-olds) who appeared to accrue no greater success as demonstration score increased in the transient information condition. All other age groups displayed a positive relationship between task score and demonstration score across all cue conditions and stages, as did even the very youngest children when cues were visible.

The findings reported regarding Stage C, which provides the best comparison to the 3 x 3 treasure chest arrays used by Wilks et al. (2021), are broadly consistent with the results presented by Wilks et al. (2021). With some exceptions, this also holds when considering our assessment of potential for ratcheting. When cues were transient, only the oldest children displayed potential for cumulative culture on Stage C, with assigned ratcheting levels steadily increasing across age brackets.

However, whereas all age groups achieved Level 3 when cues were visible on Wilks et al. (2021) Find-the-Treasure task, no age group reliably met Level 3 (taking both stages into account) on our PFR Task in the visible cues condition. Only the 2- to 3-year-olds showed a marked improvement on visible compared to transient cues, corresponding to Wilks et al. (2021) and our predictions. The classifications of the three older age groups resulted in a noteworthy departure from the results of Wilks et al. (2021) when cues were visible, driving less marked variation between cue condition in the current study.

The unexpected underperformance found in the visible cues condition may be understood in relation to two main factors. Firstly, the older children may have found the “task” to be implausibly easy when cues were visible, and thus sought an alternative solution, or instead chose to explore the grid despite knowing this would sacrifice their score. Similar performances, (i.e., in the opposite direction to predictions based on cognitive load, and most notably in participants expected to display greatest competence) have been identified in other touchscreen implementations of PFR-type Tasks in our own research group (e.g., Wilks et al., under revision). In contrast, the physical instantiation of the task that was employed by Wilks et al. (2021) may have reinforced the task contingencies by using treasure chests that contained physical ‘treasure’ and others that were materially empty. This may contrast with the electronic implementation used in the current study. Here, there was potentially less assurance (to a naïve participant) that the reward structure was 100% constant between demonstration trial and test trial, in the way that was clear with the physical objects. Secondly, our sample sizes were also considerably lower than Wilks et al. (2021). This factor may have meant that random fluctuations in the data unduly infuenced the values we used to assess potential for ratcheting.

The outcome of our simulated transmission chains also suggested cumulative effects were stronger with age, and with minimisation of memory load (with the exception of the 5-year-olds on Stage C, for the same reasons discussed above). Of particular note, the contrasting results across cue conditions in the 2- to 3-year-olds suggest that information storage/manipulation constrained the potential for cumulative culture in this group. Thus, the capacity for ratcheting using the PFR Task appears to vary both by population characteristics and contextual features.

Our simulations also highlighted some variation in the potential for cumulative culture found between the task stages. In absolute terms, the top scores from Stage C and D chains were similar, however, Stage D chains fell shorter of the *relative* optimum than Stage C (the relative optima for Stage C = 3, and Stage D = 4). Increasing the number of distractor stimuli and the number of selections to be made therfore also appears to have affected children’s ability to approach optimum performance.

Additionally, Stage D appeared to offer greater sensitivity in detecting differences between the cue types on the simulations in our younger age brackets. It is plausible that the slightly more graded reward structure on Stage D caused an increase in difficulty. Potentially, the benefits afforded by providing visible cues became more prominent on Stage D as it was more challenging. This may have enhanced the ability to identify differences in potential for ratcheting between our cue conditions. Together with the variation between stages in highest classification level reached on visible cues, inclusion of Stage D may also be beneficial when using the PFR Method with non-human primates.

To assess PFR Level 3 according to our continuum in this study we used qualitative inspection of the data only. Because the sample size was so small, this approach was more prudent as it constituted a better comparison to Wilks et al. (2021), as the differences in statistical power would have generated a divergent pattern of statistical significance. However, analyses that test whether higher scoring demonstrations had been *significantly* met could be performed on larger datasets using the current instantiation of the PFR Method.

Although we found that the older age brackets’ potential for ratcheting was more limited than we had expected in the visible cues condition, this likely resulted from over-thinking the task. The youngest age bracket, however, achieved higher levels in this condition than they did in the transient condition, in line with predictions and the results of Wilks et al. (2021). Thus, it appears that minimising the burden on memory may have increased the potential for performance to ratchet up in this group. The younger children's performance may be particularly relevant for non-human primates, suggesting providing visible cues may increase the likelihood of displaying potential for cumulative culture in these populations.

The pattern of results from the 2- to 3-year-olds demonstrates the value of the PFR Method. Although this group performed above chance overall on transient cues (demonstrating some vicarious learning), by exposing them to a range of demonstrations of varying quality, our method had the sensitivity to discern that this would likely not support cumulative build-up over episodes of transmission. This is clear from the minimal variation in success found across the range of demonstration scores. The higher classifications assigned on the visible cues condition demonstrate that this failure to accumulate benefits could not be attributed to other factors such as a general lack of motivation, or a limited understanding of the goal of the task. Thus, these results suggest that it does not always follow that an ability to selectively use information to some degree (such as proficiency on the WSLS task used in Chapter 2) would necessarily generate a ratchet effect following exposure to increasingly high-quality information.

Overall, we broadly validated our methodology against the implementation of the PFR Method developed by Wilks et al. (2021), motivating its use with non-human primate populations. As somewhat differential results were obtained between the array sizes utilised here, both Stage C and D may be useful in identifying capacity for cumulative culture in non-human primates. Additionally, in line with the aim to create optimal conditions for displaying potential for ratcheting, initial exposure to the PFR Task involving visible cues would likely bolster their success.

Chapter 4 (Well-Trained) Capuchins Have the Capacity for Cumulative Culture in a Reduced Memory Load Context.

Contributions

Donna Kean, Christine Caldwell, Elizabeth Renner and Mark Atkinson contributed to study conception, design and material preparation. The code for the task was written by Mark Atkinson. Data collection was performed by Donna Kean, Elizabeth Renner, Chloë India Wright and Amy Derrick. The data analysis was performed by Donna Kean with contributions from Christine Caldwell. The chapter was written by Donna Kean and edited by Christine Caldwell.

Introduction

The capacity of non-human species to engage in cumulative culture is fundamental to understanding the uniqueness of human culture. Capacity for cumulative culture is likely to involve the ability to augment culturally inherited information with information that has been learned through personal experience. In this manner, improvements can be made to the functionality of cultural traits over generations, equipping later generations with higher quality information than earlier ones. This process appears to be widespread in humans, but very rare in non-human populations.

Evidence from observational and experimental studies suggest non-human animals (henceforth 'animals') may possess some level of potential for cumulative cultural evolution (or 'ratcheting') in very specific contexts, also described in Chapter 1. One of the best-known examples comes from comparisons of the migratory propensities of free-ranging bighorn sheep (*Ovis canadensis*) and moose (*Alces alces*) populations that had been translocated to a novel habitat (Jesmer et al., 2018). Accumulation of knowledge regarding locations of the highest quality forage was inferred by comparing the propensities of populations that had been resident for varying lengths of time. The authors found that increased migration was positively related to duration of time since translocation. The key role of cultural learning was implicated as newly installed groups engaged in very low rates of migration, which gradually increased as populations of increasing longevity were considered. Moreover, the length of time over which migration appeared to increase would have necessarily spanned multiple generations. This implies that inherited mechanisms are unlikely to be primarily responsible. Overall, these results suggest that later generations may have benefitted from exposure to their (more knowledgeable) predecessors relative to earlier generations, a key characteristic of cumulative culture.

Another candidate comes from an empirical study of homing pigeons (*Columba livia*) using generational replacement methods (Sasaki & Biro, 2017). Subjects in an experimental transmission chain condition were found to accumulate increasingly efficient flight routes over learner generations, an effect not found in matched control conditions where individual birds and fixed pairs were exposed to the same task. As with Jesmer et al. (2018), a benefit of intergenerational learning appears to have been observed.

A more contentious example involves the use of tool sets in natural conditions by chimpanzees (*Pan troglodytes*). Tool set use is defined as the sequential use of more than one type of tool for a single purpose (Brewer & McGrew, 1990). Chimpanzees have been found to use composite tools to gather honey, and to fish for ants and termites (Boesch et al., 2009; Sanz et al., 2004, 2009, 2010). Although these behaviours has been suggested as a prime candidate to have developed by means of cumulative culture (Marshall-Pescini & Whiten, 2008), others have questioned whether the relative simplicity of this behaviour means that it is potentially inventable at the individual level, therefore not the outcome of multiple generations of social transmission (Tennie et al., 2009).

Although the examples mentioned potentially fulfil Mesoudi and Thornton's (2018) core criteria[2] for cumulative culture (detailed in Chapter 1, section '*The Cumulative Culture Paradox*'), they appear constrained in contextual range. Whereas humans seem unrestricted in their ability to improve upon the solutions of others across domains and contexts, the best non-human examples appear tightly linked to natural tendencies under specific circumstances; there is little evidence that the aforementioned non-human capacities for ratcheting could be generalised. Thus, although cumulative culture appears not to be precluded in animals, it is likely that it occurs only under very specific conditions.

Methodologically, the tasks of Jesmer et al. (2018) and Sasaki and Biro (2017) in particular could not be presented to species that do not share the necessary characteristics (i.e., migration, flight), nor the tendencies for herding/flocking which were responsible for (presumably largely incidental) transmission of beneficial modifications.

These issues were largely mitigated, however, in two related touchscreen experiments conducted with a single population of Guinea baboons (*Papio papio*), both of which claim to demonstrate cumulative culture (Claidière et al., 2014; Saldana et al., 2019). An abstract, touchscreen stimulus selection task (similar to the task used in Chapter 3) was used, the design of which was clearly outside the realm of problems this species may encounter in natural conditions. The first experiment required subjects to repeat selections that they were exposed to during demonstrations (Claidière

---

[2] Core criteria for cumulative culture i) introduction of a novel behaviour, ii) social transmission of this behaviour, iii) the behaviour must deliver a benefit in relation to existing behavioural variants, and iv) iteration of this process generating a succession of improvements.

et al., 2014). The second study reversed the rewarded strategy, instead requiring avoidance of the demonstrated selections (Saldana et al., 2019). These tasks were abstract, simple and touchscreen-based, therefore meeting the competencies of a relatively broad range of species.

Although both studies discovered a cultural evolution effect whereby responses changed over learner generations to become increasingly easier to copy (or avoid), these effects are fundamentally different to cumulative culture as we have defined it. Indiscriminately copying or avoiding vicarious information would not be a particularly effective strategy in real world scenarios. Indeed, theoretical analyses and empirical work have shown that such an approach is likely maladaptive (e.g., Giraldeau et al., 2002; Miu et al., 2020). It is entirely plausible that flexibly switching between copy and avoid strategies concurrently may be more difficult than the sum of their parts (Dunstone et al., 2021), and therefore testing these abilities separately may not satisfactorily test this ability.

Indeed, a leading cognitive explanation for the quantitative discrepancy in evidence of cumulative culture between humans and animals regards the flexibility of social learning strategies (SLS) (see Chapter 1, section '*What Makes Human Culture So Unique?*'). These strategies provide guidance about *when* social learning (as opposed to individual learning) should be employed, *what* should be copied, and *who* should be copied from. For example, if unsuccessful behaviour were observed, it may be more advantageous to seek out a more proficient social model (who), or to learn from individual trial-and-error experience rather than copying (when), or to only retain certain elements of the observed behaviour that may contribute to success (what). As such, implementation of SLSs allow the selective uptake of vicarious information so that beneficial traits can be retained, and sub-optimal or deleterious variants can be avoided (Galef, 1995; Kendal et al., 2018; Laland, 2004). A simplified SLS is integral to success on the PFR (i.e., potential-for-ratcheting) Task described in the previous chapter and discussed further below.

Examples of SLSs can be observed in even very young children, and in a wide range of non-human species (Battesti et al., 2015; Bono et al., 2018; Rakoczy et al., 2010; Range et al., 2007). However, it has been suggested that humans have a unique ability to explicitly represent such strategies, allowing greater precision and specificity when learning (Heyes, 2016). This conscious awareness of SLSs may allow increased flexibility when using vicarious information and the ability to devote cognitive resources in a deliberate manner. This may, at least partially, account for the phenomenon that humans have the capacity to engage in cumulative cultural evolution in an open-ended manner, whereas animals appear contextually restricted.

Chapter 2 and the results of Renner et al. (2021) suggested that tufted capuchin monkeys (*Sapajus apella*) have the ability to engage in a win-stay, lose-shift (WSLS) vicarious learning strategy at a basic level. The term vicarious is used here to refer to learning of information that has not been produced by the learner; specifically *social* interaction/understanding is not the focus in the Chapters 3-5, but the ability to selectively use vicarious information. The touchscreen presented PFR Task (henceforth, 'PFR Task') of Chapter 3 extended this paradigm, which involved a graded scale of solution success, making it possible to infer the outcome of linear transmission chains and thus assess capacity for ratcheting. As described in Chapter 3, extrapolation of performance at a basic level (i.e., the WSLS) to one involving more complex task demands (i.e., Stages C and D of the PFR Task) may not reflect synergistic challenges associated with applying the correct strategy in more complex scenarios. Chapter 3 confirmed that the PFR Task, which we ultimately aimed to use with the non-human primates, was a valid assessment of ratcheting ability as results were comparable to a logically similar 3D version of the task employed by Wilks et al. (2021).

The current chapter thus aimed to test two populations of tufted capuchin monkeys using the PFR Task described in Chapter 3 to assess their cumulative cultural ability, the overarching goal of this thesis. Ultimately, we hoped that the simple design of the PFR Task would create a context which, in principle, may allow a ratchet effect to be inferred using the PFR Method (outlined in Chapter 1 and 3) in non-human primates.

In contrast to Claidière et al. (2014) and Saldana et al. (2019), optimal PFR Task responding requires flexible use of both copy and avoid strategies within individual problems. In order to use the information provided by the demonstrations optimally, reliable copy-rewarded and avoid-unrewarded responding is necessary. Only this type of performance would allow the demonstrations to be improved upon (on average), indicating the capacity for ratcheting. Also contrasting with the task used with baboons, the PFR Method involves exposing participants to the full range of possible task scores and using this data to infer the performance of a transmission chain. This approach reduces the logistical challenges of running actual transmission chains (Caldwell et al., 2020) which was necessary for the two baboon studies, as the goal was to track changes to response patterns.

Capuchin monkeys may be a good candidate for displaying capacity for ratcheting for several reasons. In natural conditions, they display foraging traditions with exceptional longevity, with potential cultural variation between or within groups (Falótico et al., 2019). Arbitrary social conventions or traditions have also been documented in wild populations, involving costly signals that give an indication of group commitment (Perry, 2011; Perry et al., 2003). Experimental work has demonstrated capuchins' ability to maintain seeded artificial foraging techniques, consistent with the development of traditions in natural circumstances (Dindo et al., 2008, 2009). Tufted capuchins have also potentially displayed some ability for meta-cognitive monitoring (Smith et al., 2020), a potentially important ability underlying cumulative culture, discussed in Chapter 1. Together with the ability of some proficient individuals on the WSLS task, (Chapter 2; Renner et al., 2021), the evidence from the natural behaviour and empirical studies of capuchins suggests that some potential for cumulative culture may be displayed using the PFR Method, particularly given optimal conditions.

Given the above evidence, it is perhaps surprising that there has been no test (that we are aware of) of the potential for ratcheting over transmission events in capuchin monkeys. However, this may be attributable to the considerable logistical challenges of constructing transmission chains with non-human participants, described in Chapter 1. The current methodology was designed to address these issues, and thus presents the first assessment of the capacity for cumulative culture in capuchin monkeys.

Despite slightly unexpected findings on the visible cues condition (where demonstrated cues remained visible during test trials) in older children in Chapter 3, we predict that this context may best identify ratcheting in the monkeys. We believe the most likely explanation for this finding is that older children considered the task improbably easy, which was not expected to affect non-human primate populations. Thus, it was anticipated that providing visible cues would likely bolster the potential for cumulative improvement in non-human primates due to reduced cognitive load, as it did in the very young children.

We aimed to subsequently expose the capuchins to the transient cues condition (where the cues disappeared after the demonstration, and were not available during test trials), which required the use of working/short-term memory for optimal information use, once proficiency on the visible condition had been achieved. This was to ascertain whether the ability to use the vicariously provided information may be influenced by cognitive load, discussed further in the Transient Cues sections below.

The children's performance on the PFR Task also suggested that Stages C and D may elucidate differential ratcheting potential. Therefore we also planned to use both stages in order to provide a range of conditions that may facilitate display of this ability. However, the monkeys were initially exposed to Stages A and B in order to scaffold the learning of focal Stages C and D, where cumulative culture ability can be inferred.

Predictions

We predicted that there would be greater repetition of rewarded information than unrewarded, consistent with the optimal task strategy. It was also predicted that task score would improve with demonstration value. We were also interested in any variation between the two populations of capuchin monkeys across each of these measures, and in their potential for ratcheting according to the PFR Classification System and following simulated transmission. Results from the visible cues condition will first be reported, followed by transient cues (Kyoto group only).

Visible Cues Method

Subjects & Site

There were twenty-five subjects from two populations of capuchin monkeys based at the Living Links to Human Evolution Centre at RZSS Edinburgh Zoo and the Graduate School of Letters at Kyoto University. Subjects were never food or water deprived, and there was no experimental restraint, so subjects were always free to interact (or not) with the experimental apparatus during testing.

Ethical approval was granted by the Animal Welfare and Ethical Review Body (AWERB) of the University of Stirling, refs: (18 19) 093 and (18 19) 092. The research was also approved by Edinburgh Zoo and Kyoto University.

Edinburgh

Nineteen capuchin monkeys housed at Edinburgh Zoo (mean age at start of the experiment = 8.89 years, SD = 3.25; seven females) were involved in this study. All population and housing details for the Edinburgh monkeys were the same as Chapter 2, with the addition of two infant capuchins to the East group.

Kyoto

Six capuchins housed at Kyoto University (mean age = 15.33; SD = 6.44; three females) also participated in this study. The capuchins were housed in an indoor cage matrix of approximately 3m^2 and were fed a daily diet of primate pellets, fruit, eggs and vegetables. Water was available ad libitum. Housing and care adhered to Kyoto University Primate Research Institute's Guide for the Care and Use of Laboratory Primates (2nd ed.). Interaction with the experimental apparatus was always voluntary.

Materials

The training and experimental materials were written on PsychoPy 1.84.2 (Peirce et al., 2019) using Python2.

Edinburgh

The testing facilities were as described in Chapter 2. The task was run and presented on an ELO 1939L touchscreen monitor connected to a Microsoft Surface tablet. An Apeman camera was used to film sessions. Raisins, peanuts and sunflower seeds were used as food rewards.

Kyoto

Subjects were transported to the research room using a transport box (measuring 40 x 37 x 30 cm).

The research room had three identical research stations allowing testing of up to three subjects concurrently. Each station involved a built-to-order personal computer (CPU: Intel, Core (TM) i7-8700CPU, 3.20 GHz) connected to a touchscreen monitor (ARTS-017N-04, Minato Holdings Inc., Tokyo, Japan) for running the task. The monitor was integrated into one wall of a transparent acrylic test box (measuring 46 x 46 x 52 cm) allowing access to the entire touchscreen interface. The opposite wall of the test box was removable which allowed the subject to be transferred from the transport box into the test box. Two levers were attached below each touchscreen monitor (one to the left and one to the right to account for variation in handedness) for activating problems. White noise was played to mask any external sounds and partitions were used to separate individual research stations to ensure subjects had no visual access to other parts of the room. Sessions were filmed using Sony (HDR-CX670) video cameras. Sweet potato and apple pieces were used as food rewards and were delivered by a universal feeder (Biomedica BFU310-P100).

Procedure

Edinburgh

Access to the research cubicles for the Edinburgh capuchins was as described in Chapter 2 (see ‘Procedure’, first paragraph). The researcher activated each problem (for training and test). Food rewards were provided manually.

Kyoto

Subjects entered the transport box voluntarily and were carried to one of three research stations in the research room and transferred to the test box. A food reward was manually provided at the start and conclusion of each session. The subject activated each problem (training and test) by depressing the lever. Food rewards were dispensed automatically.

Pre-Task Touchscreen Training

The pre-task training served to (re)introduce the subjects to general touchscreen interaction procedures. The first three stages of the programme were as outlined in Chapter 2. However, the current study added an additional fourth and final stage. This was identical to the previous stage (iii) touching small stimuli) but required subjects to select a ‘continue’ button before presentation of the small stimulus. Subjects were rewarded only after the continue button and the small stimulus had been selected. The continue button training was added as, unlike on the WSLS Task (Chapter 2), a continue button had to be touched after the demonstration trial on the PFR Task in order to proceed to the test trial. As described in Chapter 3, the continue button was added to the PFR Task to ensure subjects had full opportunity to attend to the demonstration. As in Chapter 2, the criterion to pass each stage was ≥80% over three consecutive sessions for the Edinburgh monkeys (each session comprising 12 problems), and ≥80% in a single session for the Kyoto monkeys (involving 100 problems).

Edinburgh

Subjects who had not previously completed all levels of our pre-task touchscreen training program re-entered training at the stage they had last encountered. Subjects who had previously completed our training program were required to meet criterion again over two consecutive sessions on the third stage (touching small stimuli). This served as a reintroduction to the program. Criterion then had to be met on stage four to familiarise this population with the continue button.

Kyoto

All Kyoto based subjects were previously touchscreen and lever trained. They were exposed to the fourth stage of pre-task touchscreen training only (touching continue button and small stimuli) to familiarise them with the stimulus selection nature of the study, and the continue button.

PFR Task Procedure

The PFR Task itself (i.e., the demonstration trial, continue button, test trial, and optimum strategy) (Figure 3.1) used in the current chapter was identical to the task used in Chapter 3. The only differences were to the procedure and are described in detail below.

The capuchins were exposed to each stage of the task (Table 3.1) in order of increasing complexity (from A to D). Subjects began with PFR Task pre-criterion training sessions until a criterion of accuracy was met (≥80%) over a varying number of sessions (see population specific details below). Accuracy here refers to repetition of rewarded and avoidance of unrewarded demonstration selections. At Stages A and B this was calculated per selection, i.e., whether the correct strategy was used for each individual selection within problems. At Stages C and D, this was calculated per problem, and one selection error (i.e., a single repetition of an unrewarded, or failure to re-select a rewarded stimulus) was allowed per problem for the problem to be recorded as successful overall.

Meeting criterion was followed by a set number of post-criterion problems on the same stage before progressing to the subsequent task stage. The number of post-criterion problems for each stage (Table 4.1) was determined by the associated number of levels of demonstration, i.e., the number of possible demonstration scores. We aimed to conduct at least twenty problems for each demonstration level per monkey. This was to ensure enough power for the individual level analyses described below. Therefore, the number of post-criterion problems completed per stage was (at least) the number of demonstration levels multiplied by twenty. For example, on Stage C there were four demonstration levels (*0, 1, 2,* and *3*), therefore the number of post-criterion problems required overall was 80 (4*20). However, some individuals completed more than this minimum number of post-criterion problems if, for example, they finished Stage D much earlier than the other individuals in the group. The task was programmed to display an equal number of demonstrations of each value per session, randomly ordered, at training and test.

Table 4.1. Number of post-criterion problems for each PFR Task stage.

Stage	**A**	**B**	**C**	**D**
Number of Problems	40	60	80	100

For both populations, food rewards were distributed after all selections on a given trial had been made. Food reward structures (detailed below) were designed to be similar to what each population had been exposed to (and were therefore accustomed to) in previous experiments.

All data collection took place from April 2019 - March 2020. Population specific procedural details will now be outlined.

Edinburgh

The Edinburgh based capuchins were exposed to a varying number of problems per session depending on task stage (A = 24, B = 21, C = 20, and D = 20 problems) for both PFR Task training and post-criterion sessions. This was done to ensure an equal number of problems involving each demonstration level, and to avoid exceeding the maximum number of rewards allowed per session as set by the rules of the facility (24). Subjects were required to meet our performance criterion over three consecutive sessions to pass a given stage and progress to the post-criterion sessions.

On Stage A (where 1 target was available) the Edinburgh subjects received half a food reward if the target was found (i.e., half of one raisin or peanut). On Stages B to D (2, 3 and 4 targets available, respectively), a graded food reward structure was used:

- Half a food reward was given if at least one target but less than all targets was found.
- Two halves were delivered if 100% of the targets were found.

However, during the course of the study, this structure was changed for Stages C and D due to visibly reduced motivation from the first monkeys to reach these stages. We modified the structure by alternating between giving one half and two halves for finding 2/3 (Stage C) and 2/4 or 3/4 (Stage D) targets. For example, on Stage D, if on Trial 1 a monkey found 2/4 targets, they received one half; if on Trial 2 they also found 2/4 targets they would receive two halves; and if on Trial 3 they got 3/4 targets, they were given one half, and so on. This increased the likelihood of receiving the maximum reward on a given trial.

If four unrewarded stimuli were selected in succession, subjects received a short time-out and a sunflower seed to maintain motivation. Although sunflower seeds were desirable to the subjects, the task rewards were relatively more desirable.

Kyoto

Each test session for the Kyoto capuchins involved 60 problems on both PFR Task training and post-criterion sessions, except for the first test session of Stage A where each monkey was exposed to 100 problems. The number of problems was reduced after this first session in order to provide larger (and more motivating) food rewards without exceeding the monkey's daily food intake. On Stage A subjects were required to meet the criterion level of accuracy over two consecutive sessions to pass. On Stages B-D meeting criterion was required on a single session to pass. This variation was due to the decreased probability of meeting criterion by chance after Stage A.

On Stage A, a food reward was given if the target was found and no reward was given if no target was found. Food rewards were provided using a probabilistic reward structure for Stages B-D. This was necessary rather than a simpler graded rewarded structure (as described for the Edinburgh capuchins) as the automatic reward dispenser could not efficiently dispense more than one food item per trial.

- **Stage B**: finding one target resulted in a 50% chance and finding two targets had a 100% chance of receiving a reward.
- **Stage C**: finding one target had a 50% chance, two targets had a 75% chance, and finding all three targets had a 100% chance of receiving a reward.
- **Stage D**: locating one target had a 50% chance, two targets had a 66% chance, three targets had an 82% chance, and four targets had a 100% chance of food reward.

Analysis

Descriptive statistics only were computed for the PFR Task training data (i.e., pre-criterion problems) from Stages A-D.

The following analyses were all performed on the Stage C and D post-criterion data only. Two subjects (Alba and Ximo) completed post-criterion trials for Stage C but not Stage D.

As in Chapter 3, four general linear mixed models (GLMMs) were constructed to test our hypotheses regarding repetition and task score on the PFR Task, using the *glmer* function of the *lme4* package (Bates et al., 2014) on R (R Core Team, 2020). Models 1 and 3 analysed Stage C (9-stimuli) post-criterion data, and models 2 and 4 analysed the Stage D (16-stimuli) post-criterion data.

Models 1 and 2 included *repeats* as a binary outcome variable: repetition of any stimulus chosen during the demonstration trial was categorised as a 'repeat' [1] and selecting any alternative stimulus was recorded as a 'shift' [0]. Information type (rewarded vs unrewarded) and group/population (Kyoto vs Edinburgh) were included as fixed effects, as was their interaction.

Models 3 and 4 included *task score* as a binary outcome measure where any selection that was rewarded, i.e., found a target, was recorded as successful [1] and unrewarded selections were recorded as unsuccessful [0]. Demonstration score (0-3 on Stage C; 0-4 on Stage D), group, and their interaction were included as fixed effects.

If included in a model, information type and group were sum coded so that their levels were -1 and 1 and included as random slopes. Participant ID was included as a random intercept in all models to account for the repeated measures nature of the study. As described in Chapter 2 '*Analysis*' section, a maximal effects structure was used for all GLMMs. No singular fit or convergence issues were encountered.

Estimated marginal means were calculated to test the direction of any significant interactions using the package *emmeans* (Lenth et al., 2019) on R, and are given on the log odds ratio scale. For post-hoc tests involving demonstration score, comparisons were made using the upper and lower quantiles.

To assess whether each population and individual subject met the criterion for Levels 1 and 3 according to our Classification System, we took the mean task score for each stage overall, and for demonstration scores 1-2 for Stage C, and 1-3 for Stage D, as in Chapter 3. One sample t-tests were calculated to assess whether chance, and each demonstration score was outperformed. Pearson's correlations were also performed to test for the presence of a linear relationship between demonstration score (1-3 for Stage C, and 1-4 for Stage D) and task score, and therefore whether Level 2 had been met.

For each population and subject, linear transmission chains were also simulated for each stage, as described in Chapter 3 '*Results*' section.

Visible Cues Results

PFR Task Training to Criterion

Figure 4.1 displays the proportion of repetition of rewarded and unrewarded demonstration trial selections over the test trials (i.e., before criterion was met) and displays significant inter-individual variation. The number of PFR Task training problems for each subject and whether they met criterion on each stage can be found in Appendix 3.

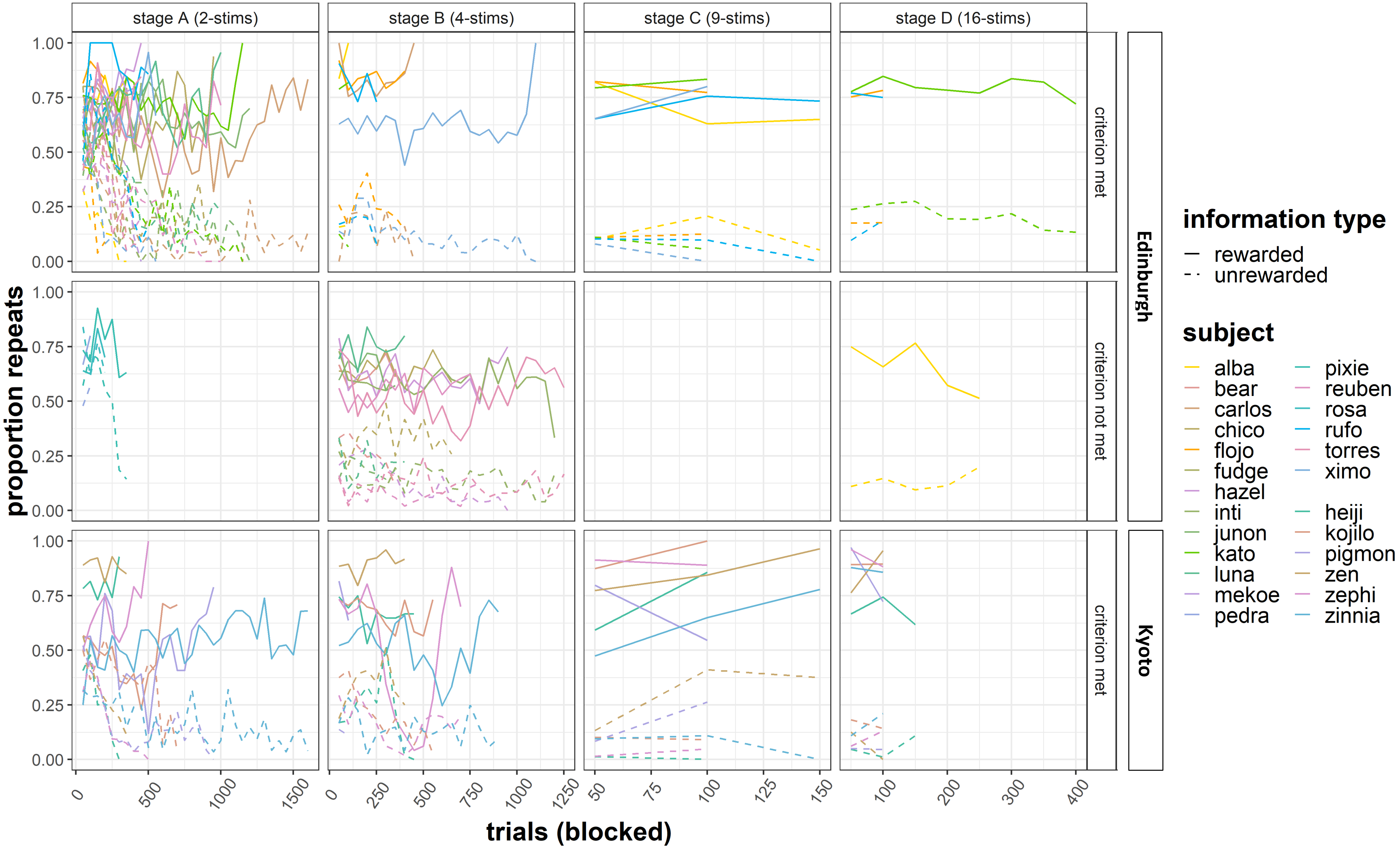


Figure 4.1. Proportion of repeats over test trials (in blocks of 50) split by information type and faceted by stage and population. *NB: X-axes are on different scales across stages, particularly at Stage C. All Kyoto monkeys met criterion for each stage.*

Eleven subjects completed the set number of post-criterion problems on Stage C, and nine on Stage D. This data only will be considered for the remainder of the Results section.

Repeats GLMMs

Stage C: 9-stimuli

The GLMM was significantly more predictive than the null equivalent (χ^2 (3) = 30.64, *p* < .001) (Figure 4.2). Information type was significant (*b* = 1.69, SE = 0.14, z = 11.88, *p* < .001) indicating more repeats when information was rewarded compared to unrewarded. Group was not a significant predictor of repetition (*p* = .92).

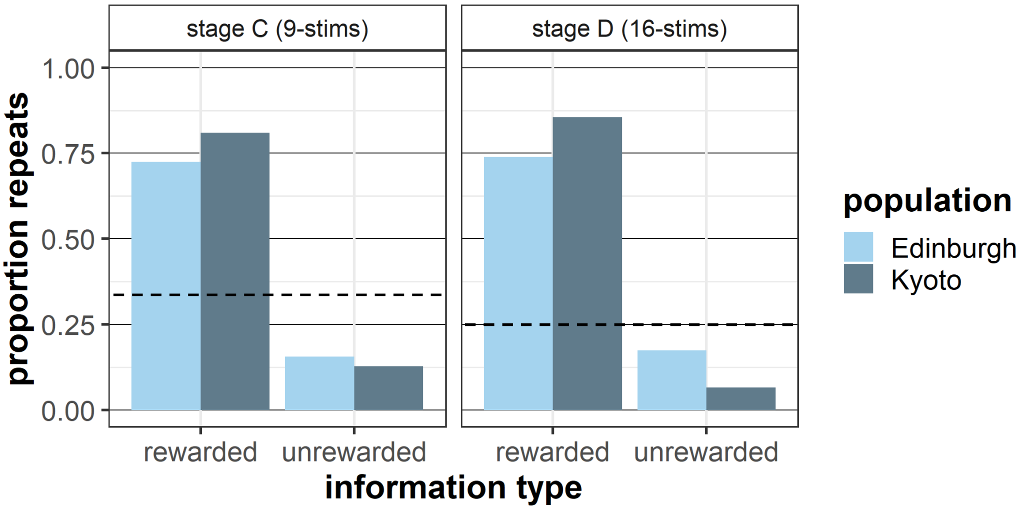


Figure 4.2. Average repetition by information type and population, faceted by stage. *Dashed lines indicate chance of repetition (33% on Stage C; 25% on Stage D).*

The interaction between information type and group was also significant (*b* = 0.28, SE = 0.14, z = 2.01, *p* = .04). Post-hoc results indicated that the effect of information type was stronger in the Kyoto group (*b* = 3.94, SE = 0.38, z = 10.42, *p* < .001) than Edinburgh (*b* = 2.81, SE = 0.42, z = 6.67, *p* = <.001) (Figure 4.3).

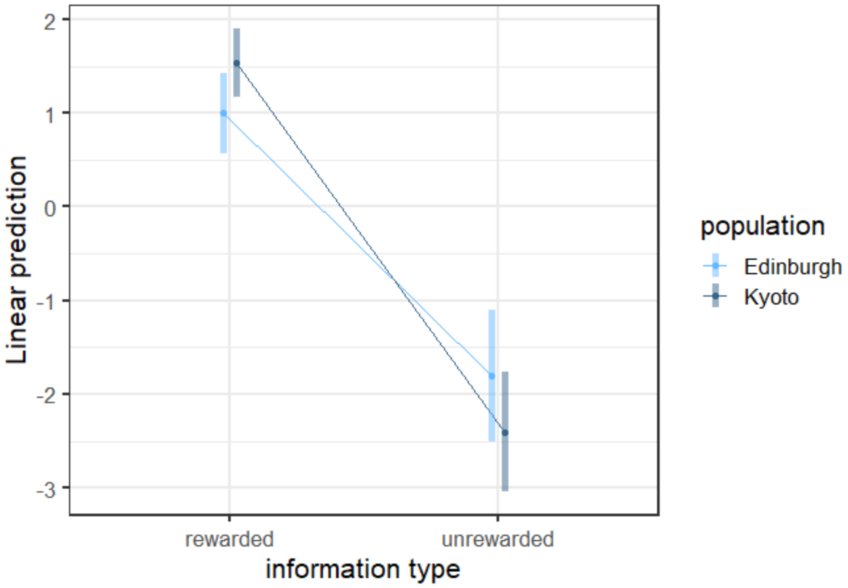


Figure 4.3. Interaction between population and information type on Stage C.

Stage D: 16-stimuli

For Stage D, the model was again a better fit of the data than the null model (χ^2 (3) = 43.86, *p* < .001) (Figure 4.2). Again information type was found to significantly predict repetition (*b* = 1.79, SE = 0.06, z = 27.90, *p* < .001), and group was non-significant (*p* = .93).

As with the Stage C model, the interaction between information type and group was significant (*b* = 0.51, SE = 0.06, z = 7.88, *p* < .001). Again, information type appears to have a stronger effect on the Kyoto (*b* = 4.60, SE = 0.15, z = 31.52, *p* <.001) compared to Edinburgh (*b* = 2.57, SE = 0.21, z = 12.16, *p* < .001) monkeys (Figure 4.4).

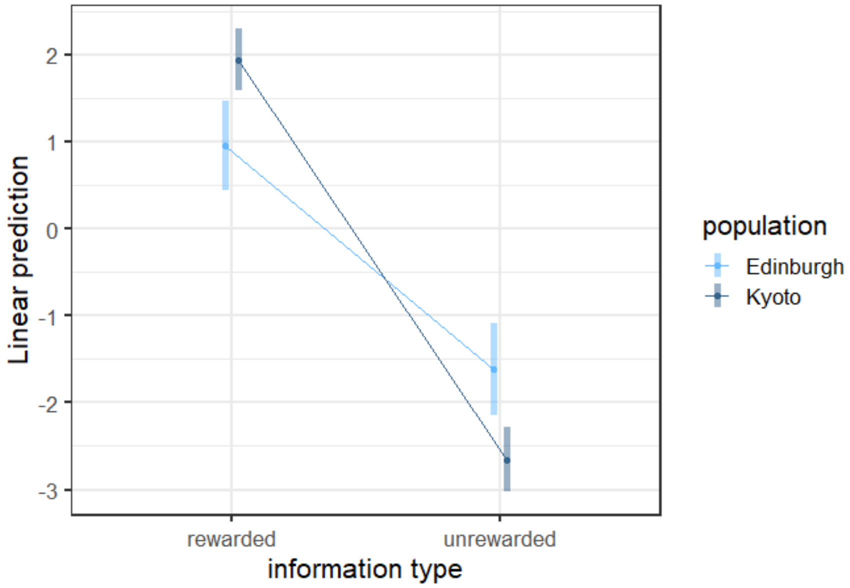


Figure 4.4. Interaction of population and information type on Stage D.

Task Score GLMMs

Stage C: 9-stimuli

The Stage C task score GLMM was better than its null equivalent (χ^2 (3) = 26.67, *p* < .001) (Figure 4.5). Task score significantly increased with increasing demonstration score (*b* = 0.43, SE = 0.05, 9.26, *p* < .001). Group was not a significant main effect (*p* = .38).

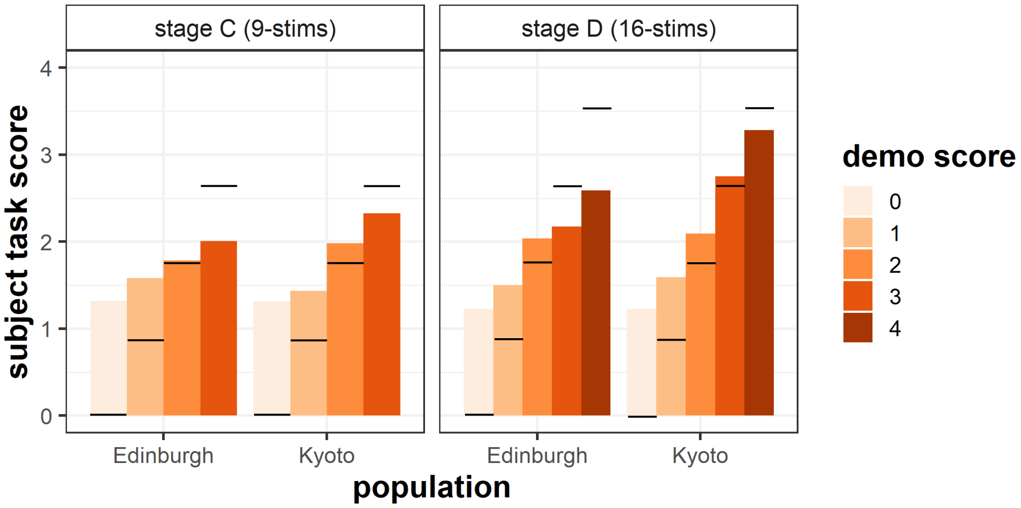


***Figure 4.5. Mean task score (number of targets found) by population and demonstration score, faceted by stage.** Horizontal black lines highlight the associated demonstration score level.*

The interaction between demonstration score and group had a significant effect on task score (*b* = 0.11, SE = 0.05, z = 2.34, *p* = .02). The results from post-hoc testing revealed that demonstration score had a slightly stronger effect on the number of targets found by the Kyoto group (*b* = 0.54, SE = 0.06, z = 9.43, *p* <.001) compared to Edinburgh (*b* = 0.32, SE = 0.07, z = 4.42, *p* <.001) (Figure 4.6).

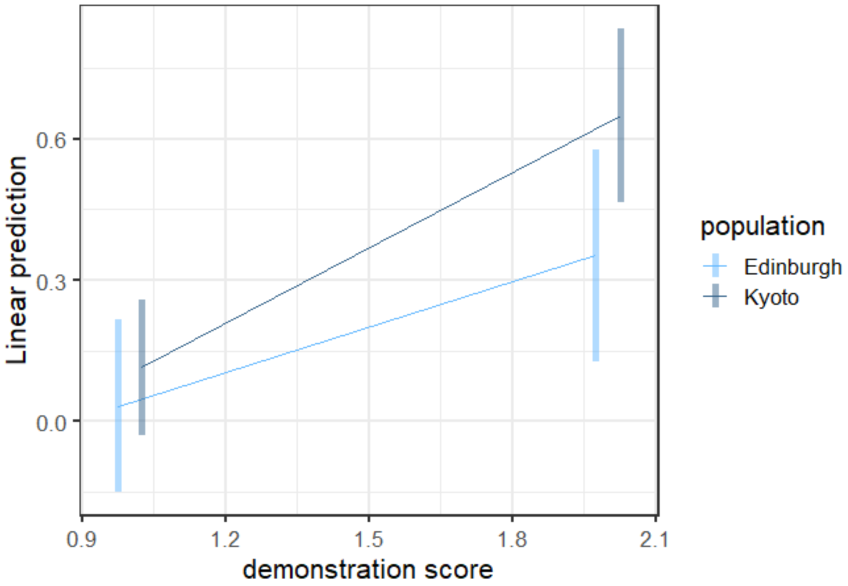


Figure 4.6. Interaction of population and demonstration score on Stage C.

Stage D: 16-stimuli

Compared to the null model, the Stage D task score model was more predictive of the data (χ^2 (3) = 1267.61, *p* < .001) (Figure 4.5). As above, task score improved with increasing demonstration score (*b* = 0.46, SE = 0.02, z = 22.71, *p* < .001) and group was not a significant predictor of task score (*p* = .88).

Consistent with Stage C, the interaction between demonstration score and group was significant (*b* = 0.08, SE = 0.02, z = 3.87, *p* < .001) and a post-hoc test indicated this was also in the same direction: a stronger effect of demonstration score was found in the Kyoto (*b* = 1.08, SE = 0.03, z = 31.91, *p* < .001) compared to the Edinburgh (*b* = 0.77, SE = 0.07, z = 10.36, *p* < .001) capuchins (Figure 4.7).

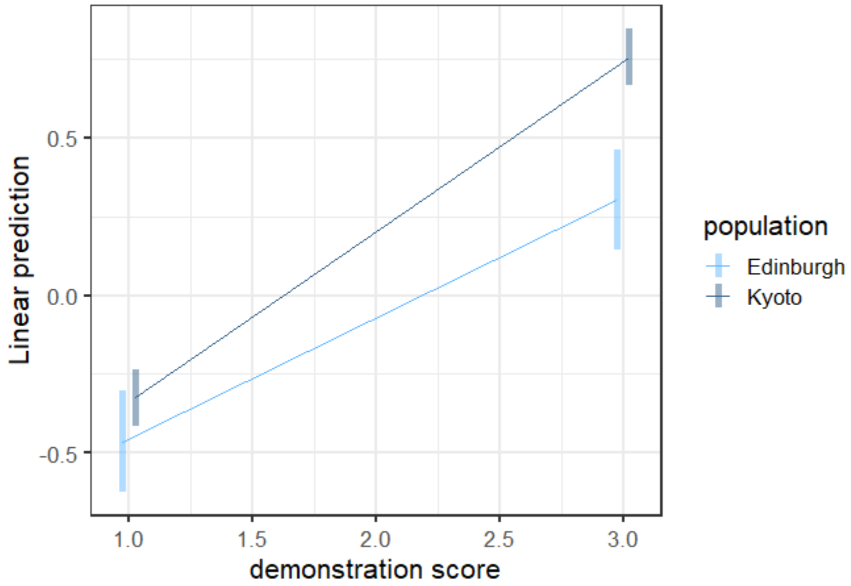


Figure 4.7. Interaction of population and demonstration score on Stage D.

Potential for Ratcheting

This section will now outline the PFR Classifications assigned, and the results of the simulated transmission chains.

The mean task scores, t-tests and correlations that were used to assign PFR Classifications can be found in Appendix 4 for each population (also Figure 4.5) and subject (also Figure 4.8). For assessing whether PFR Classification Level 3 was met, we inspected both whether there was a numerical improvement relative to the associated demonstration score, as well as whether this difference was significant. The latter was possible in the current chapter due to the larger quantities of data (relative to Chapter 3, where significance tests were not performed for Level 3). Significance tests were performed for Levels 1 and 2 (as in Chapter 3) as otherwise there was a 50% chance of scores falling on the right side of the criterion even if simply selecting at random.

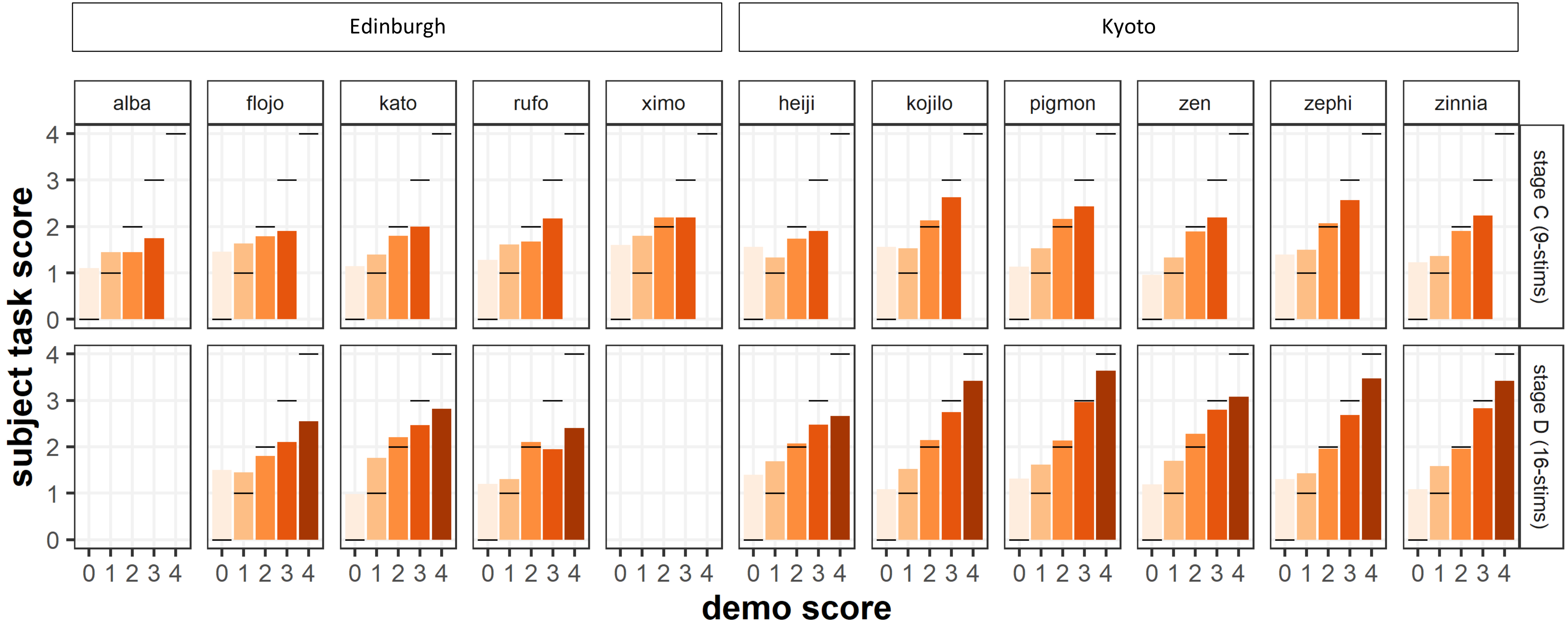


__Figure 4.8. Mean task score (number of targets found) by demonstration score, faceted by subject and stage.__ Horizontal black lines indicate the associated demonstration score level.

Figure 4.9 displays the highest classification level met by each group and subject on Stage C (9-stimuli) and D (16-stimuli). Kyoto as a group and two individual monkeys (Kato and Zen) met our criteria for Level 3, displaying potential for ratcheting with significance on Stage D. The Edinburgh group and several other individuals across both stages also achieved this level, although their scores were not significantly above the threshold value. The lowest Level met was 1, and only by two individuals of the Edinburgh group (Alba and Flojo) and only on Stage C.

It appears that Stage D may have been more sensitive to detecting potential for ratcheting than Stage C, as higher levels were generally reached by both groups, and at the individual level.

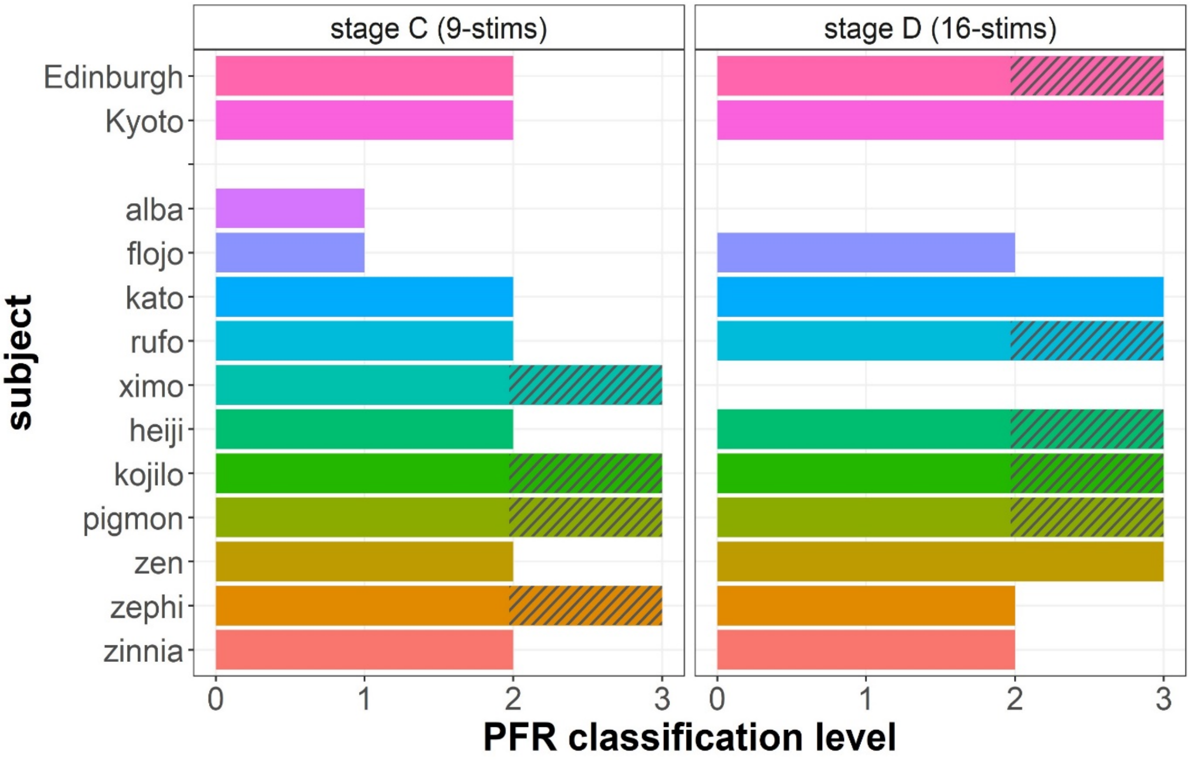


Figure 4.9. Top PFR Classification level met by each subject for both Stage C and D. *Shaded sections indicate that Level 3 was met numerically only (without significance).*

Linear transmission chains were also simulated to infer performance on the PFR Task across repeated transmission events. The Kyoto group reached slightly higher top scores than did the Edinburgh group overall (Figure 4.10). At the individual level, there was substantial individual variation between the top scores reached (Figure 4.11). Overall, however, even the highest scoring performances plateaued at well below the optimum level, which represents fully consistent application of the correct strategy. These results suggest that in a transmission chain scenario, some accumulation of success may be possible, but it is likely that this progress would be relatively limited, whereby only a few transmission events may result in improvement, whilst some would result in actual deterioration in performance. Figure 4.12 displays sub-optimal information use across the range of possible demonstration scores within both groups and stages. Especially in the Edinburgh population and on Stage D, there appears to be a slight decrease in rates of repetition in response to rewarded information in particular. This suggests that as the number of rewarded stimuli increased, application of the optimal strategy in response to this information type (repetition) decreased.

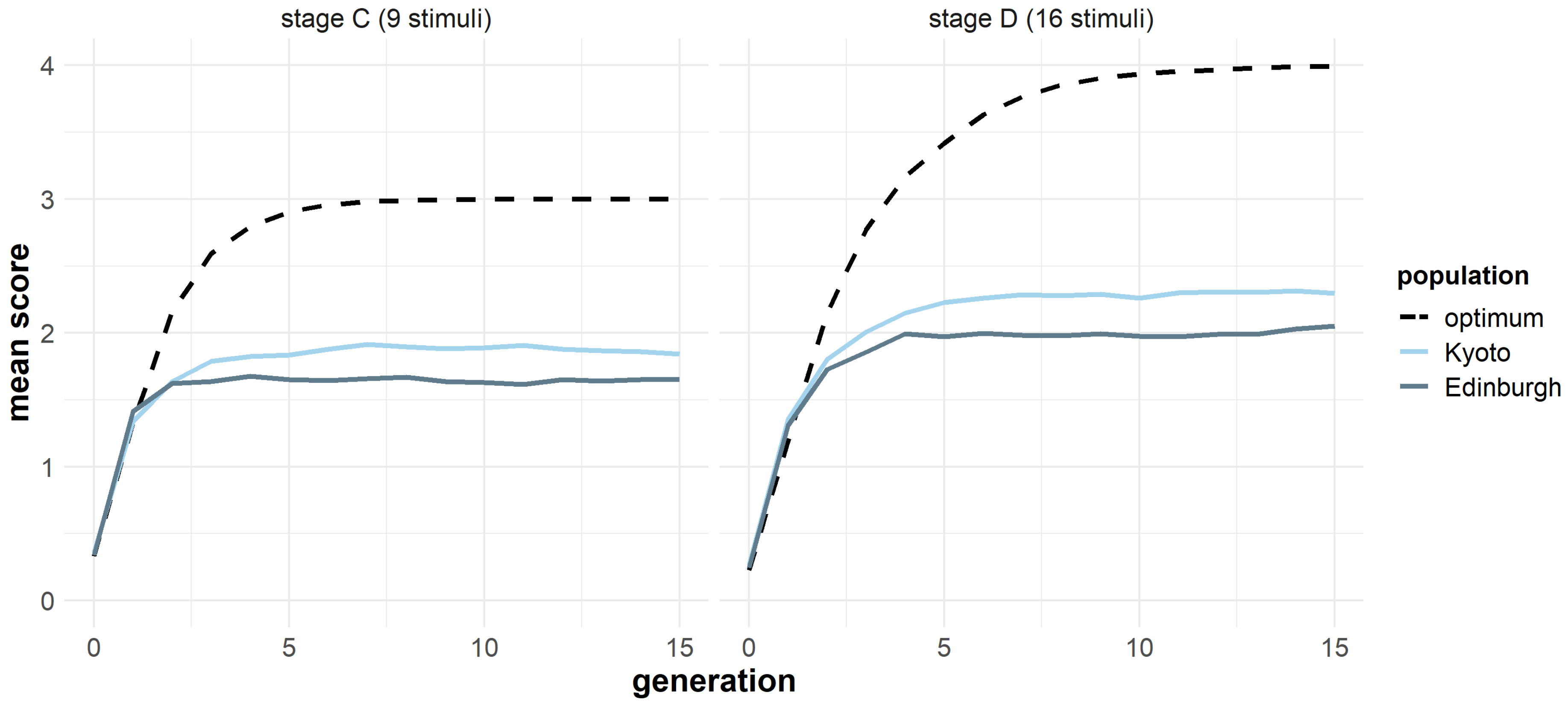


*Figure 4.10. **Simulated transmission chain performance for each population.** Simulations from both stages are displayed. The dashed line represents optimal performance, i.e., consistent application of the correct strategy. Chancel level for Stage C was 1/3 and for Stage D was 1/4.*

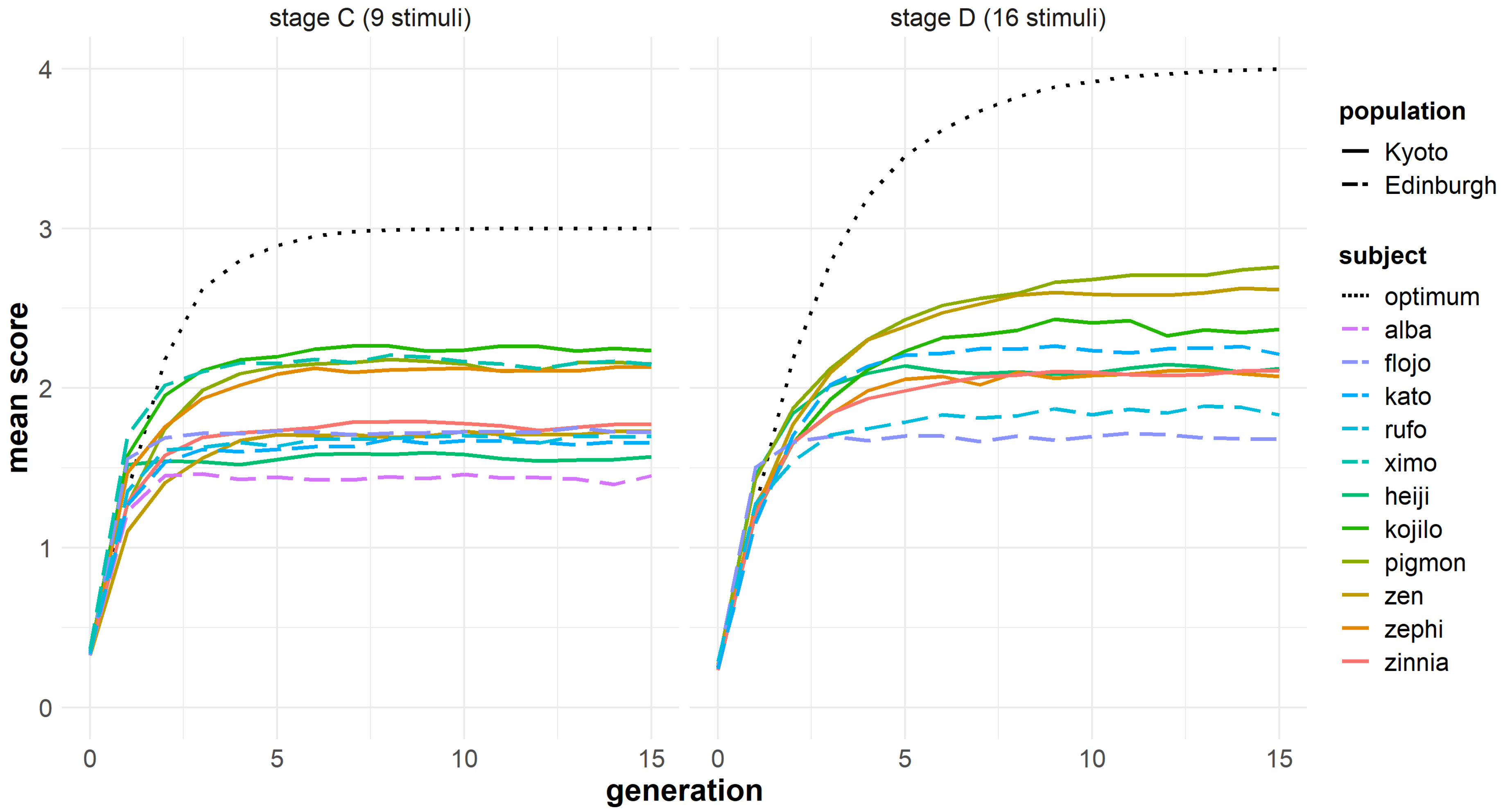


Figure 4.11. Simulated transmission chain performance for each subject. *Simulations from both stages are displayed. The black dotted line represents optimal performance i.e., consistent application of correct strategy. Chancel level for Stage C was 1/3 and for Stage D was 1/4.*

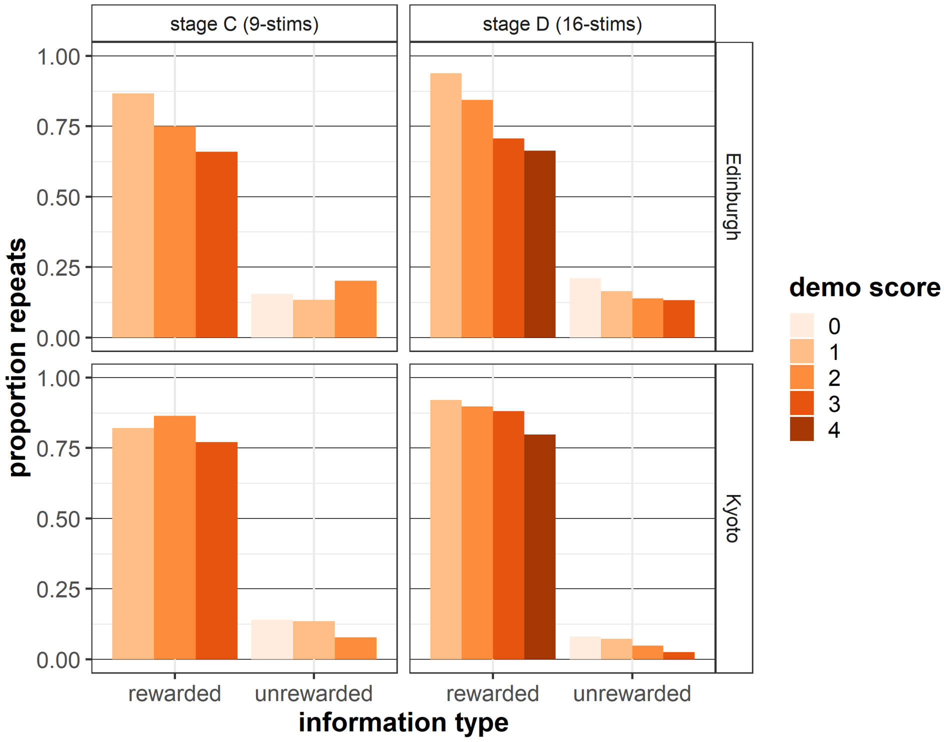


Figure 4.12. Proportion of repetition by information type and demonstration score, faceted by stage and group.

Visible Cues Discussion

In the context of low memory load, both capuchin populations displayed the potential for cumulative culture on Stage D at the group level, as did several individuals across both stages. Moreover, the performance of the Kyoto monkeys as a group and two individual monkeys was significantly above the associated criterion for this level of classification. These findings suggest that, given favourable conditions, the type of discriminatory learning that may underlie ratcheting is available to this species in a manner that may support cumulative improvement over multiple generations.

However, the simulations highlight that the cumulative progression suggested by this classification may be restricted, occurring over only a very limited number of transmission events. Furthermore, not all subjects progressed to the final stages of the task, therefore the limited success found also reflects only the most successful individuals. The pattern of results from two subjects (Alba and Flojo) on Stage C also displays that even some very well-trained monkeys do not display any suggestion of ratcheting, apparently gaining no additional benefit from theoretically higher-value demonstrations. As discussed in Chapter 3, this indicates that single instances of selective vicarious learning cannot be extrapolated to circumstances where the complexity of the learning environment has increased. For example, even from proficient performance on the win-stay, lose-shift task (WSLS) of Chapter 2, it cannot be assumed that this would generate a ratchet effect using the PFR Task of the current chapter, where there are more stimuli, and varying proportions of rewarded and unrewarded information revealed by the demonstration.

The population of capuchins housed at Kyoto appeared to consistently outperform the Edinburgh based capuchins, even though they had been subject to the exact same performance criterion prior to progressing to the test phase. It is likely that this is simply due to the greater experience this group has had with touchscreen-based tasks.

In order to evaluate the role of memory using the current method, we had hoped to contrast the performance on the visible cues condition with the transient cues condition (Figure 3.1, right panel), as with the children in Chapter 3. The transient condition does not provide cues to the outcome of the demonstration trial during the test trials, instead requiring the use of memory. This condition is consistent with the conditions of the WSLS task in Chapter 2. Extrapolating from the performance of the youngest children in Chapter 3, we may expect the monkeys to struggle with this condition considering their limited performance even when provided with visible information, above. However, progressing to the transient cues condition was only possible with the Kyoto group due to time constraints.

The two cue conditions were presented to the capuchins consecutively, as opposed to interleaving these conditions as in Chapter 3. This was done as we speculated that the (presumably easier) visible cues context may scaffold learning in the transient cues condition. Indeed, more monkeys met criterion on the 2-stimulus stage on visible cues compared to the almost identical WSLS task employed in Chapter 2, which involved transient cues. Although this could also have resulted from having previous experience with the paradigm, this is unlikely to fully explain the divergent results.

However, minimal progress was made by the monkeys in this section of the study, with no individual reaching focal Stages C or D, preventing appraisal of potential for ratcheting. Descriptive statistics for this limited progress will be presented.

Transient Cues Method

All details were identical to those described for the Kyoto based monkeys on visible cues except that i) the task cues were transient (Figure 3.1, right panel) and ii) descriptive statistics only were produced as no individual reached Stages C or D on the PFR Task. The transient cues condition continued directly after the visible cues condition had been completed.

Transient Cues Results

PFR Task Training to Criterion

When cues were transient, four monkeys reached criterion on Stage A, however none achieved criterion performance on Stage B. The proportion of repeats on Stages A and B are displayed in Figure 4.13. The number of PFR Task training problems completed by each individual on each stage can be found in Appendix 5.

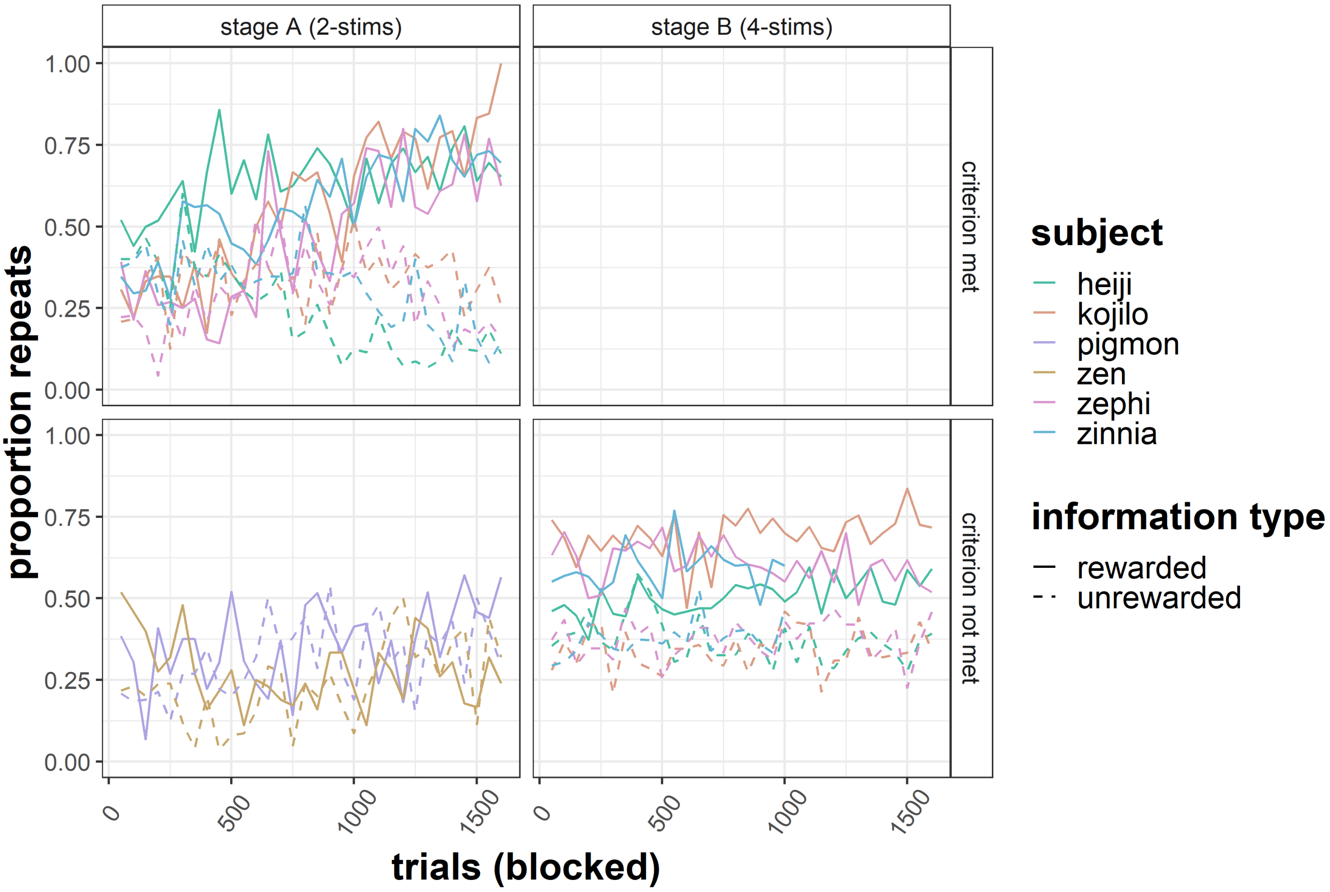


Figure 4.13. Average repeats over PFR Task training trials (in blocks of 50) on transient cues, split by information type and faceted by stage. *NB: trials were capped at 1600; see Appendix 5 for full trial numbers.*

Transient Cues Discussion

Interestingly, there was very little progress made by the Kyoto capuchins when the cues provided were transient, despite substantial exposure to the task. This was despite the inclusion of a continue button which minimised the risk of 'missing' the demonstration. Although we cannot confidently conclude the reason for this, there are a few possibilities. Firstly, optimally responding to the PFR Task may have been compounded by the cognitive load involved in memorising demonstration trial outcomes. Secondly, or potentially in addition, initial experience with visible cues may have hindered performance on transient, due to potentially confusing cues to success. For example, on the visible condition perhaps only the lasting cues themselves were used to guide responses, and the demonstration trial itself was not considered relevant 'information'. If so, then this initial experience of the task may not have facilitated performance on the transient cues condition, and could even have been detrimental to the monkeys' ability to learn this new version of the task. It is plausible that the removal of the information utilised by the subjects may have presented an insurmountable barrier.

Discussion

This study provides the first example of the capacity for cumulative culture in capuchin monkeys, apparent at both the group and the individual level. We would like to place the emphasis here on context, as we suggest that minimising cognitive load was likely instrumental in eliciting this effect, similar to findings in young children (see Chapter 3 and Wilks et al., 2021). However, this hypothesis has not been directly tested in the current study as no subject reached the necessary proficiency when memory load became necessary for success on the same task (transient cues condition). Although this may be congruent with an explanation that the additional executive burden caused this poor performance, we cannot rule out the possibility that order effects played a significant role and that the task might otherwise have been mastered.

Although ratcheting up of performance was observed, this was at relatively low levels. The outcomes of simulated transmission suggest that more precise application of the optimal strategy was required for ratcheting to approach optimal levels even though visible cues were provided. It appears that once the demonstrated performance was substantially higher than might be expected by chance, it became harder to substantiate improvement; for example, no group or individual outperformed demonstrations (on average) that scored three on Stage D. This may reflect real life scenarios, where behaviours that are the result of improvements over many generations may be more difficult to master, compared to behaviours that have resulted from the trial-and-error learning of a single individual.

This may be influenced by the chance levels of success on the current task across the range of possible demonstration scores. Consider Stage D: there is a 1-in-4 chance of repeating a demonstration selection, and a 3-in-4 chance of selecting an alternative. Thus, the probability of outperforming the demonstration is higher for lower scoring demonstrations, where lower rates of repetition (i.e., selecting an alternative) are favoured, whereas high scoring demonstrations require higher rates of repeating demonstration selections, which are less likely by chance. This effect is compounded by the reduced rate of repetition of rewarded stimuli that was found as demonstration score increased (particularly in the Edinburgh group, who displayed lower overall performance) further dampening the ratchet effect. Thus, if there is substantial imprecision in the responder's execution of the correct strategy, this will likely disproportionately affect performance following higher value demonstration trials, on average, resulting in loss of information.

In contrast, results from a study using a logically similar task with human adults found that performance closely matched optimal responding (Mackintosh, 2020). Potentially, explicit representations of learning strategies may have engendered greater precision and specificity regarding information use in human participants. This may be made possible by greater inhibitory or attentional control in humans (Beran et al., 2016).

The pattern of results from the capuchins suggests that in real life scenarios, ratcheting of cultural behaviours or artefacts may be limited in progress. This may be congruent with the existing evidence from natural populations. For example, putative examples of cumulative culture in nonhuman primates, which have included tool sets (Sanz et al., 2009), stone throwing (Kühl et al., 2016), and potato washing (Schofield et al., 2018), each involve relatively simple technologies that could potentially be invented by a single individual. Congruently, our results suggest that if cumulative culture were to be found in the natural behaviour of non-human primates, it would likely be relatively simple, involving very few transmission events that would result in only limited increases in functionality. Overall, low-precision use of information may elucidate the paradox that cultural evolution is pervasive in non-human animals, but cumulative culture is rare.

To address the limitations of previous methodologies that have elicited a ratchet effect in animals (discussed in Chapter 1) the current task sacrificed ecological validity in favour of flexibility and efficiency (returned to in Chapter 6). However, this abstraction means that generalisation of our results must be applied with caution. Although some capuchins met our criterion for ratcheting at the topmost level of classification, this involved a significant degree of training and manipulation of the learning conditions. Such advantageous learning conditions are likely to be reflected in few, if any, of the real-world challenges faced by natural populations of capuchin monkeys.

However, our results also reveal the utility of the PFR Method. Not only does it have the power to assess populations which may otherwise be untestable, due to logistical issues, but this method also differentiates a range of ratcheting potential. For example, Alba and Flojo met our performance criteria on Stage C and clearly showed some evidence of vicarious learning (by scoring above chance overall). However, it was found that this type of performance might not necessarily equate to the ability to improve upon demonstrated information over multiple transmission events, in a manner that supports cumulative culture. This more fine-grained differentiation between the ability to learn socially and the ability to accumulate beneficial modifications may better reflect real world capacities.

The current study failed to test hypotheses regarding potential variation in performance between the two cue conditions. Although a within-subjects comparison would have been ideal, the ability of other non-human primate populations on transient cues remains an interesting avenue for research and will be detailed in the next chapter.

The current chapter provides empirical evidence of the potential for cumulative culture in a non-human primate species. The findings were consistent with the view that cumulative culture is contextually dependent rather than an all or nothing phenomenon (although this could not be directly tested), whereby optimal environmental conditions may play a vital role in the expression of this ability in animals. However, the limited extent of ratcheting displayed here suggests that low precision information use may hinder the development of cumulative culture in natural conditions.

Chapter 5 (Very Well-Trained) Baboons Show Potential for Cumulative Culture in a Memory Load Context.

Contributions

This study was conducted in collaboration with a research group based at the Cognitive Psychology Laboratory (CPL) (*Laboratoire de Psychologie Cognitive*) of Aix-Marseille University. Donna Kean, Christine Caldwell, Elizabeth Renner, Mark Atkinson and Nicolas Claidière contributed to study conception, design and material preparation. The task code was written, and the data collection was implemented by the CPL group. The data analysis was performed by Donna Kean with contributions from Christine Caldwell. The chapter was written by Donna Kean and edited by Christine Caldwell.

Introduction

The results of Chapter 4 indicated that tufted capuchin monkeys (*Sapajus apella*) may have the potential for cumulative culture (or 'ratcheting') under favourable learning conditions. Reducing the memory load associated with the PFR (i.e., potential-for-ratcheting) Task by providing visible cues (demonstrated cues that remained present during test trials) may have facilitated the potential for cumulative culture displayed by this population. Indeed, the beneficial effects of reducing memory load were apparent in young children, where the presence of lasting cues allowed even the youngest children to display a ratchet effect (Chapter 3; Wilks et al., 2021); this finding was not matched in a transient cues condition (where the cues were available during demonstrations, but did not remain visible during test trials) and appeared to present a greater challenge to younger children.

The visible/transient cues distinction reflects realistic variation in the demands involved in social transmission across varying contexts, as outlined in Chapter 1 (section '*What Makes Human Culture So Unique?*'). Briefly, it has been suggested that working/short-term memory may play an important role in the emergence of cumulative cultural evolution, and that non-human primates have a more restricted working memory capacity than humans (Balter, 2010; Carruthers, 2013; Read, 2008; Wynn & Coolidge, 2011). Contexts that impose demands on the short-term storage and manipulation of socially learned information may be more challenging for certain populations, compared to contexts of low memory load. As such, potential for ratcheting may vary, or be constrained, by memory capacity.

As discussed in Chapter 4, previous experiments with a population of captive Guinea baboons (*Papio papio*) provided evidence that they could separately use transient cues regarding what to copy (Claidière et al., 2014), and what not to copy (Saldana et al., 2019) on touchscreen stimulus selection tasks. Both studies involved a grid of sixteen stimuli where four stimuli were highlighted during demonstrations, much like Stage D of the PFR Task used in the previous two chapters. However, the task goals varied whereby the former required selection (reported in Claidiere et al., 2014) (Figure 5.1 A) and the latter required avoidance (Saldana et al., 2019) (Figure 5.1 B) of the highlighted stimuli. In both cases provision of a reward depended on selection/avoidance of at least three out of the four highlighted stimuli. In both cases, the cues provided during the demonstration were transient and thus placed demands upon memory in order to use the information during the test trials. In this respect, these tasks bore similarities to the transient cues condition presented to the children and Kyoto based capuchins on the PFR Task in Chapters 3 and 4 (Figure 3.1).

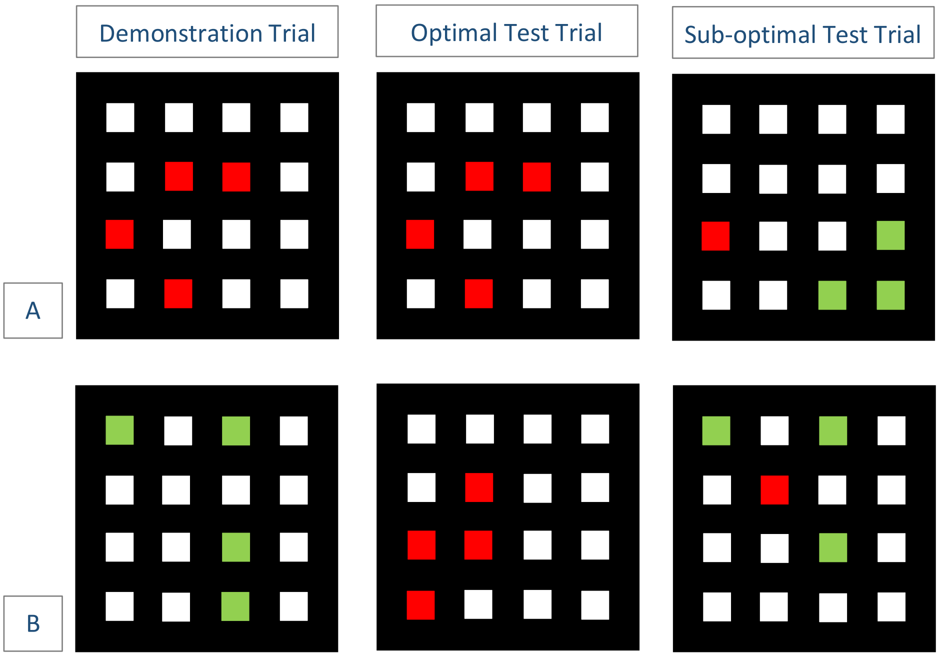


Figure 5.1. Example problems from the tasks used by A) Claidière et al. (2014) involving a copy-goal and B) Saldana et al. (2019) involving a non-copy goal. *The successful strategy was to A) repeat rewarded (red) and B) avoid unrewarded (green) stimuli. Test trial responses that follow optimal and sub-optimal strategies are displayed.*

Transmission chains were created by exposing subjects to the outputs of previous 'generations' (the individuals in a chain) in sequence i.e., the stimuli selected by the first generation in a chain became the cues presented to the second generation, and so on. Thus, when the goal was to copy, subjects were rewarded for repeating the actual selections made by previous baboons, and the non-copy goal involved avoidance of selections of previous individuals.

The authors of these studies described their findings as showing cumulative cultural evolution. However, it should be noted that the phenomena described, whilst certainly demonstrating cultural evolution in the broader sense, fall short of our own definition of cumulative culture in a critical way. Importantly, the target responses (i.e., what constituted success on these tasks) were themselves changing over repeated transmission, because there was no externally defined goal other than to copy/avoid the response of the predecessor. The effect of this was that task difficulty reduced over transmission, as the patterns of responses made by the baboons changed in the direction of typical errors. Across both tasks, this resulted in clustering of the four target stimuli, where the highlighted stimuli became increasingly grouped together and thus easier to repeat or avoid. Hence the performance improvement found over learner generations was essentially an outcome of the task itself becoming easier, as opposed to being the result of transmitted social information becoming increasingly beneficial due to an accumulation of experience on the same task.

However, the latter can emerge from responses to a task with a fixed external goal, i.e., a fixed reward structure that does not vary across generations in a transmission chain, which necessitates strategic responding, such as the PFR Task used in the previous two chapters. Episodes of transmission in such a scenario provide information that can benefit subsequent learners regarding what to copy or avoid, which can draw them closer to an externally defined optimum response (i.e., finding the target locations). This feature allows the potential for social transmission to generate increasingly effective/successful *responses*, in contrast to the increasingly easy task that develops if the reward structure is shaped by response biases.

As described in Chapter 4, mastery of a 'copy' or 'shift' rule independently may not sufficiently test the ability to employ these responses simultaneously and flexibly depending on the information provided in a particular problem. The current chapter aimed to test the same population of baboons using the PFR Method (outlined in Chapter 3) to ascertain the ability to generate a cumulative effect given the requirement of strategic responding. As this population has extensive experience with executing copy and avoid rules separately, they may be an ideal candidate for learning the contingencies of the PFR Task.

In line with the population's previous experience, the PFR Task used in Chapters 3 and 4 was adapted to exploit the superficial design features of the stimulus selection tasks already familiar to the baboons (e.g., rewarded stimuli were coloured red and unrewarded stimuli were green, Figure 5.1). As their past experience involved transient cues, this was also held constant in the current study, making the task design logically similar to the transient cues condition (see Figure 3.1, right panel) presented to the children in Chapter 3 and, with little success, to the Kyoto capuchins in Chapter 4. Although we may generally expect increased memory load to limit the potential for cumulative culture in non-human primates, our knowledge of this population's prior experience and ability to use transient cues gave us reason to believe that this mode of presentation might be sufficient for them to display some potential for cumulative culture.

Predictions

In contrast to Chapter 4, the baboon population was exposed to transient cues. However, our predictions regarding the expected pattern of results remained the same. Firstly, we predicted more repetition of rewarded compared to unrewarded information on the PFR Task. We also expected task score to increase with the value of the demonstrated information (i.e., demonstration score).

Finally, we were interested in the top PFR classifications (see Table 3.2) assigned in this population as well as performance when transmission chains were simulated based on PFR Task responses.

Method

Subjects & Site

Twenty-five Guinea baboons (*Papio papio*) were involved in this study (mean age at the start of the experiment = 10.33 years, SD = 2.33; nine females) housed at the CNRS Primate Centre in Rousset-sur-Arc. Subjects were from two separately housed groups (group 1: N = 6; group 2: N = 19). The indoor enclosures for group 1 and group 2 measured 4 x 3m and 4 x 6m, and their outdoor enclosures measured 5 x 8m and 25 x 30m respectively. Climbing structures were provided in the enclosure for behavioural enrichment, and the monkeys were fed once daily on a mixture of monkey chow, fruit and vegetables. Water was available ad libitum and subjects were never food deprived.

The data were collected in an automated, open testing environment where subjects had free access to testing booths which were located adjacent to their outdoor enclosures (full details can be found in Fagot et al., 2014, 2015). All subjects had a microchip implant for identification purposes, and all research participation was voluntary.

This research was carried out in accordance with French standards and received approval from the national French ethics committee, the "Comité d'Ethique CE-14 pour l'Expérimentation Animale" (approval number APAFIS#2717-2015111708173794). Procedures were also consistent with the guidelines of the Association for the Study of Animal Behaviour.

Materials

The task was written on E-Prime (version 2.0.10.242). Subjects were tested in automated, computerised testing booths involving a freely accessible test chamber, a 19-inch EloTouch touchscreen connected to a Dell computer, a food dispenser and a radio frequency identification (RFID) reader. The touchscreen was accessible from the test chamber by reaching through one of two hand ports and could be viewed through a viewing port (measuring 7 x 7cm).

Sessions were recorded on a Sony video camera, and dry wheat grains were provided as food rewards.

Procedure

The baboons were free to engage with the task at any time, and there was no limit to the number of problems they participated in per session. Sessions began when a subject presented their arm to the RFID reader which provided identification. The system stored progress data for each individual allowing automatic presentation of the appropriate stage to each subject.

The PFR Task design used here was logically similar to that of the PFR Task used in Chapters 3 and 4, based on the design outlined by Caldwell et al. (2020). As the current experiment was also similar to the design of previous studies with the same population (see Claidière et al., 2014 and Saldana et al., 2019), the subjects were experienced with the protocol. The key procedural details, and the main differences to the previous two chapters will be outlined here.

The stages of the task used with the baboons and the order of their presentation were identical to Chapter 4 (Table 3.1). Each stage of the current task also involved a matrix of white, square-shaped stimuli (Figure 5.2). However, unlike the previous PFR Task, on Stages A-C the stimuli were always presented in the same central grid locations, and the background colour was always black. This was because the baboons were experienced in this type of task, and therefore no development of location biases was anticipated with this population.

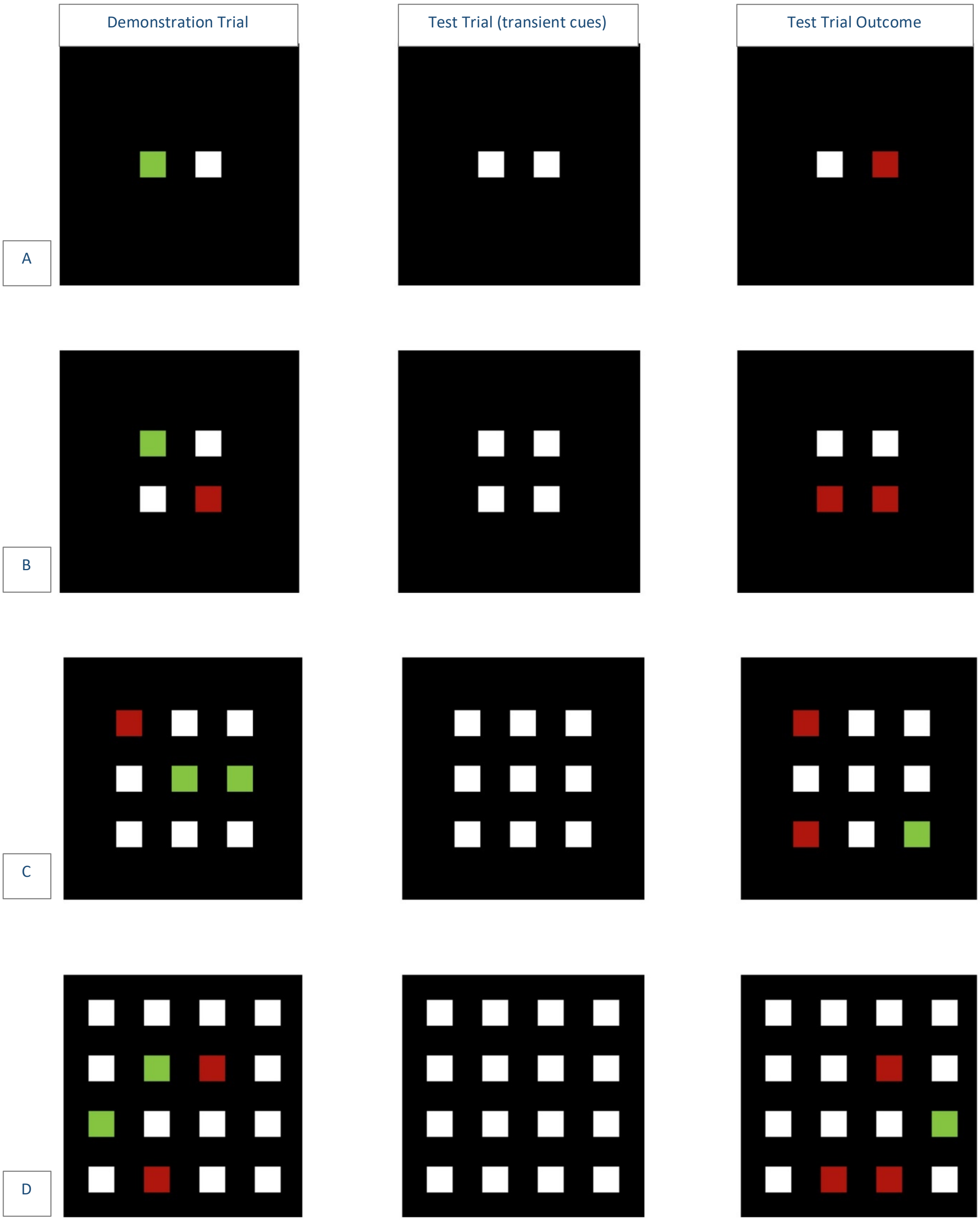


Figure 5.2. Example problems from each stage (A-D) of the baboon PFR Task. *Columns from left to right: demonstration trial; the screen subjects made their selections from during the test trial (transient cues condition); example test trial outcome after subjects had made all selections (displaying optimal strategy use).*

As in previous chapters, each problem involved a demonstration trial followed by a single test trial, and the optimum strategy was also the same (repeat rewarded and avoid unrewarded demonstrated selections). However, during the demonstration trial for the baboons, no animated agent made the selections in succession. Instead, the entire set of demonstrated information (Figure 5.2) was presented concurrently for 400 milliseconds; this was the same as in the previous studies with this population (Claidière et al., 2014; Saldana et al., 2019). Rewarded stimuli were coloured red and those that were unrewarded were green. As in Chapters 3 and 4, the location of the rewarded and unrewarded stimuli during demonstration trials was randomly assigned, and there was an equal number of trials of each demonstration score per block (e.g., on Stage D, an equal number of demonstrations scoring *0, 1, 2, 3* and *4*). If the subject touched the screen during the demonstration trial, the problem was aborted, and the same demonstration was re-presented.

The test trial screen (Figure 5.2) was then presented for the baboons to make their selections. Subjects had three seconds to make each individual selection, otherwise the problem was aborted and presented again, i.e., the same demonstration was presented followed by the test trial. In Claidière et al. (2014) and Saldana et al. (2019) stimuli selected by the subjects became black when touched, however in the current study they turned red or green (but with 50% transparency), depending on whether it was a target square. This was modified to reinforce the task contingencies, as shift responses in the current task could each result in one of two outcomes (target found or not), which was not the case in the previous two tasks. If three or more selections that did not reveal a target were made in succession, the entire screen turned green and there was a brief timeout (3 seconds). There was a four second inter-problem interval before the next problem was automatically activated.

PFR Task pre-criterion training blocks involved 60 problems, and also involved an equal number of presentations of each possible demonstration score. As the subjects were free to engage with the task at any time and had control over when to finish a session, blocks had the potential to run over multiple sessions. As in Chapter 4, PFR Task training problems were presented until a criterion of accuracy (≥80%) was met. Accuracy here refers to the strategy of re-selecting all revealed targets and avoiding all revealed non-targets within a problem. On Stages C and D, one error (i.e., failure to re-select a cued target, or re-selection of a cued non-target) was permitted per problem for it to be recorded as correct (consistent with Chapter 4). The criterion level of accuracy was required over two blocks for Stage A (2-stimuli) and on one block for the subsequent stages (B-D; 4- to 16-stimuli), as with the Kyoto capuchins. As noted above, blocks could continue over multiple sessions, therefore there was the potential for criterion to be met across separate sessions. The probabilistic reward schedule for the baboons was also identical to that of the Kyoto capuchins:

- **Stage A**: reward given if target found, and no reward if no target was found.
- **Stage B**: finding one target resulted in a 50% chance and finding two targets had a 100% chance of receiving a reward.
- **Stage C**: finding one target had a 50% chance, two targets had a 75% chance and finding all three targets had a 100% chance of receiving a reward.
- **Stage D**: locating one target had a 50% chance, two targets had a 66% chance, three targets had an 82% chance, and four targets had a 100% chance of food reward.

The PFR Task training problems were followed by a set number of post-criterion problems. The number of required post-criterion problems was the same as Chapter 4 (Table 4.1), and also contained an equal number of presentations of each possible demonstration score.

Analysis

Descriptive statistics only were calculated for the PFR Task training data (i.e., pre-criterion training problems).

The four GLMMs constructed here were identical to Chapter 4 except that group/population was not included in any model, and therefore there were no interaction terms. Briefly, the first two models included repetition (i.e., a repeat of any demonstration trial selection) as the binary outcome variable and involved the Stage C (9-stimuli) and D (16-stimuli) data, respectively. Information type (rewarded vs unrewarded) was included as a fixed effect and a random slope. The third and fourth models analysed the Stage C and D data, respectively, and included task score (i.e., whether a target was found or not) as the binary outcome measure. Demonstration score (0-3 on Stage C; 0-4 on Stage D) was included as a fixed effect and a random intercept. Participant ID was included as a random intercept in all four models.

As described in the Chapter 2 '*Analysis*' section, a maximal effects structure was used. Demonstration score was dropped as a random slope effect from both task score models due to singular fit. All models were computed using the *glmer* function of the *lme4* package (Bates et al., 2014) on R (R Core Team, 2020).

The analyses performed to assess Potential for Ratcheting were identical to Chapter 4, but will be described briefly here. Mean task scores (overall and separated by demonstration score) were calculated for each group and individual. One sample t-tests were conducted to assess the criterion to meet PFR Classification Levels 1 and 3 (Table 3.2), and Pearson's correlations tested the Level 2 criteria. Finally, simulations of linear transmission chains were conducted for the group overall and for each individual baboon for both Stages C and D (see Chapter 3 for a full description of how the simulations were constructed).

Results

PFR Task Training to Criterion

Appendix 6 displays the number of PFR Task training problems for each subject. The proportion of repetition of both information types on all stages can be found in Figure 5.3. One individual on Stage B and five individuals at Stage D completed PFR Task training trials but did not meet criterion; this data is displayed separately in Figure 5.3 (middle row). Unfortunately, none of the data from five individuals, and the Stage D data from one individual could not be included in any inferential analyses. A technical glitch allowed these subjects to progress to post-criterion blocks without having met criterion on the PFR Task training problems. Descriptive statistics for these monkeys are included in Appendix 6 and Figure 5.3 for transparency, but their data was not considered for the remainder of the Results section.

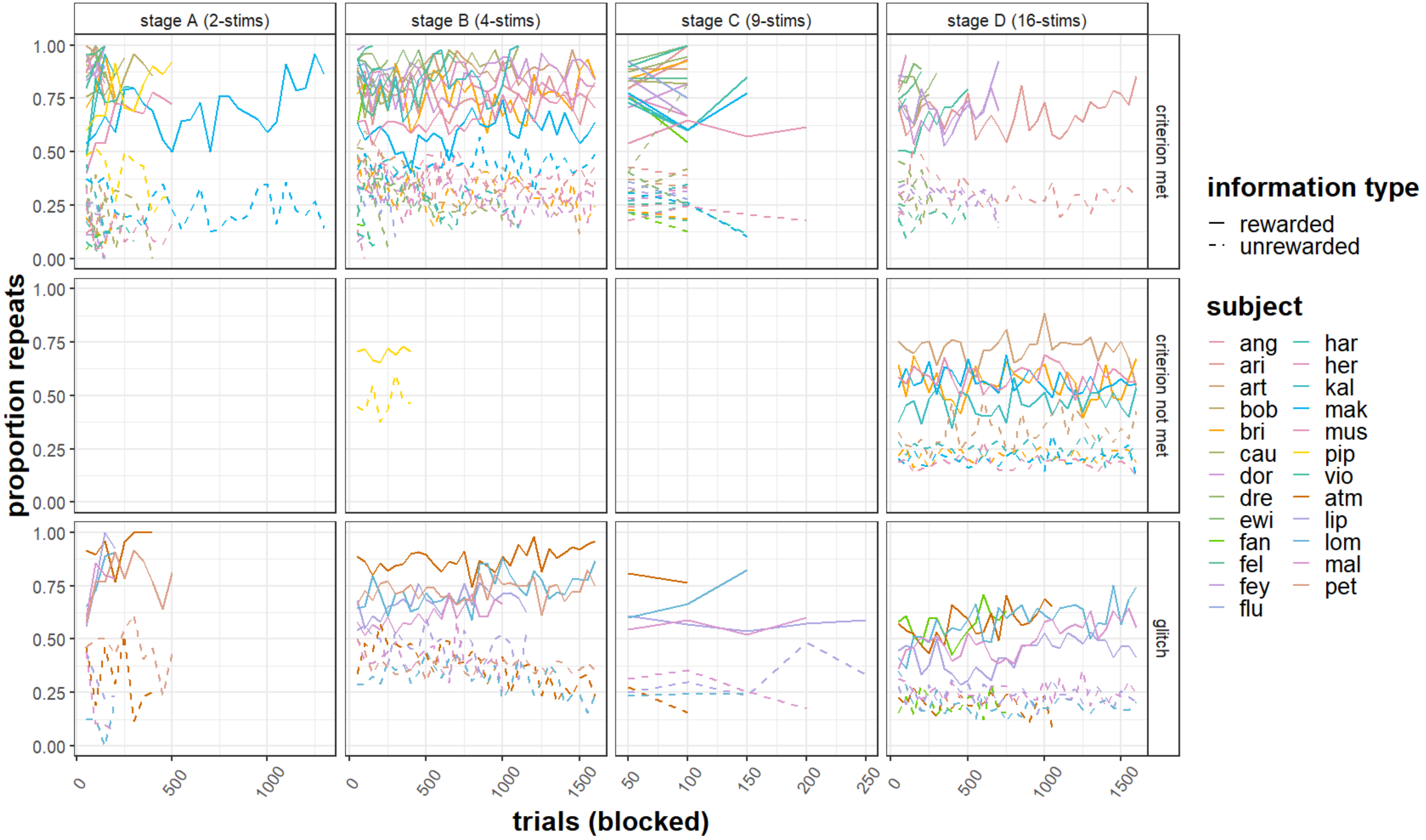


Figure 5.3. Proportion of repeats over PFR Task training trials, grouped in blocks of 50 and separated by information type. *Faceted by stage, and whether criterion was met. The performance of those affected by computer glitch (atm, fan [Stage D only], lip, lom, mal, pet) are also included for transparency. NB: the x-axes are on different scales, particularly on Stage C. Trials were capped at 1600.*

Eighteen of the baboons unaffected by the technical glitch completed the post-criterion problems for Stage C, and twelve completed both Stage C and D. This data only will be considered for the remainder of the Results section.

Repeats GLMMs

Stage C: 9-stimuli

The Stage C model was found to be more predictive of the data than the null equivalent (χ^2 (1) = 52.12, *p* < .001) (Figure 5.4). Information type was found to be a significant predictor (*b* = 1.22, SE = 0.07, z = 17.24, *p* < .001) with rewarded information associated with increased repeats in comparison to unrewarded.

Stage D: 16-stimuli

Again the repeats GLMM for Stage D was found to be a better fit of the data than the null model (χ^2 (1) = 33.80, *p* < .001) and information type was significant (*b* = 1.13, SE = 0.08, z = 13.06, *p* < .001) (Figure 5.4).

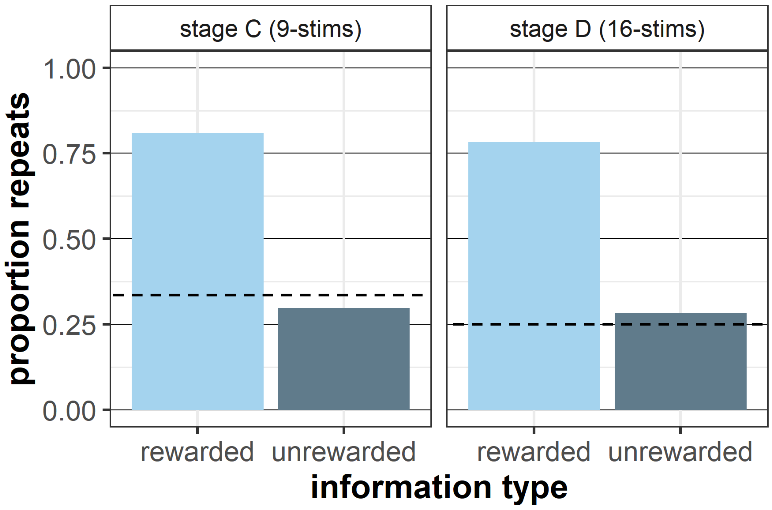


Figure 5.4 Average repetition by information type and stage. *Dashed lines indicate chance of repetition (33% on Stage C; 25% on Stage D).*

Task Score GLMMs

Stage C: 9-stimuli

The Stage C model with task score as the outcome measure was found to be a better fit of the data than the null model (χ^2 (1) = 530.10, *p* <.001) (Figure 5.5). Increasing demonstration score was significantly associated with greater task score (*b* = 0.66, SE = 0.03, z = 21.80, *p* < .001).

Stage D: 16-stimuli

Again, the Stage D model was found to be significantly more explanatory than the null equivalent (χ^2 (1) = 670.65, *p* < .001) (Figure 5.5) and demonstration score was significant (*b* = 0.57, SE = 0.02, z = 24.29, *p* < .001).

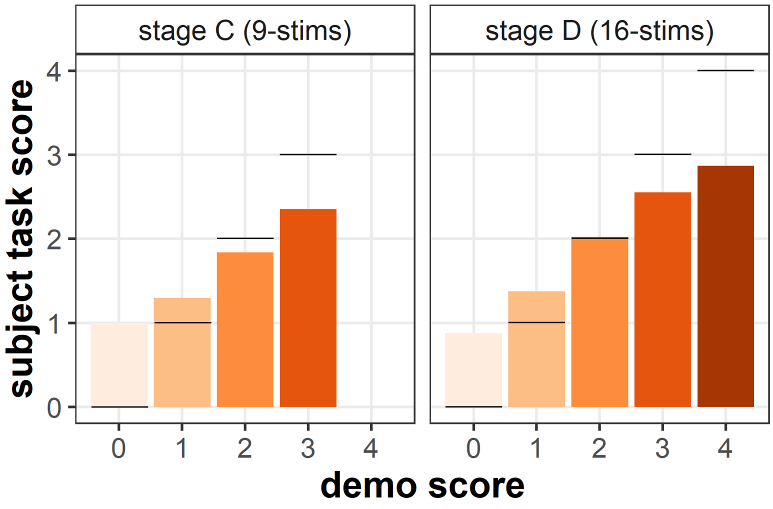


Figure 5.5. Mean task score (number of targets found) by demonstration score, faceted by stage. *Horizontal black lines highlight the associated demonstration score level.*

Potential for Ratcheting

Capacities for cumulative culture in this population will now be outlined. As for the capuchins in Chapter 4, the performance for the group and each individual was classified according to our PFR Classification System (Table 3.2), and their responses were used to simulate linear transmission chains. Mean task scores, and the results from the t-tests and correlations can be found in Appendix 7 for the group overall (Figure 5.5) and for each individual (Figure 5.6).

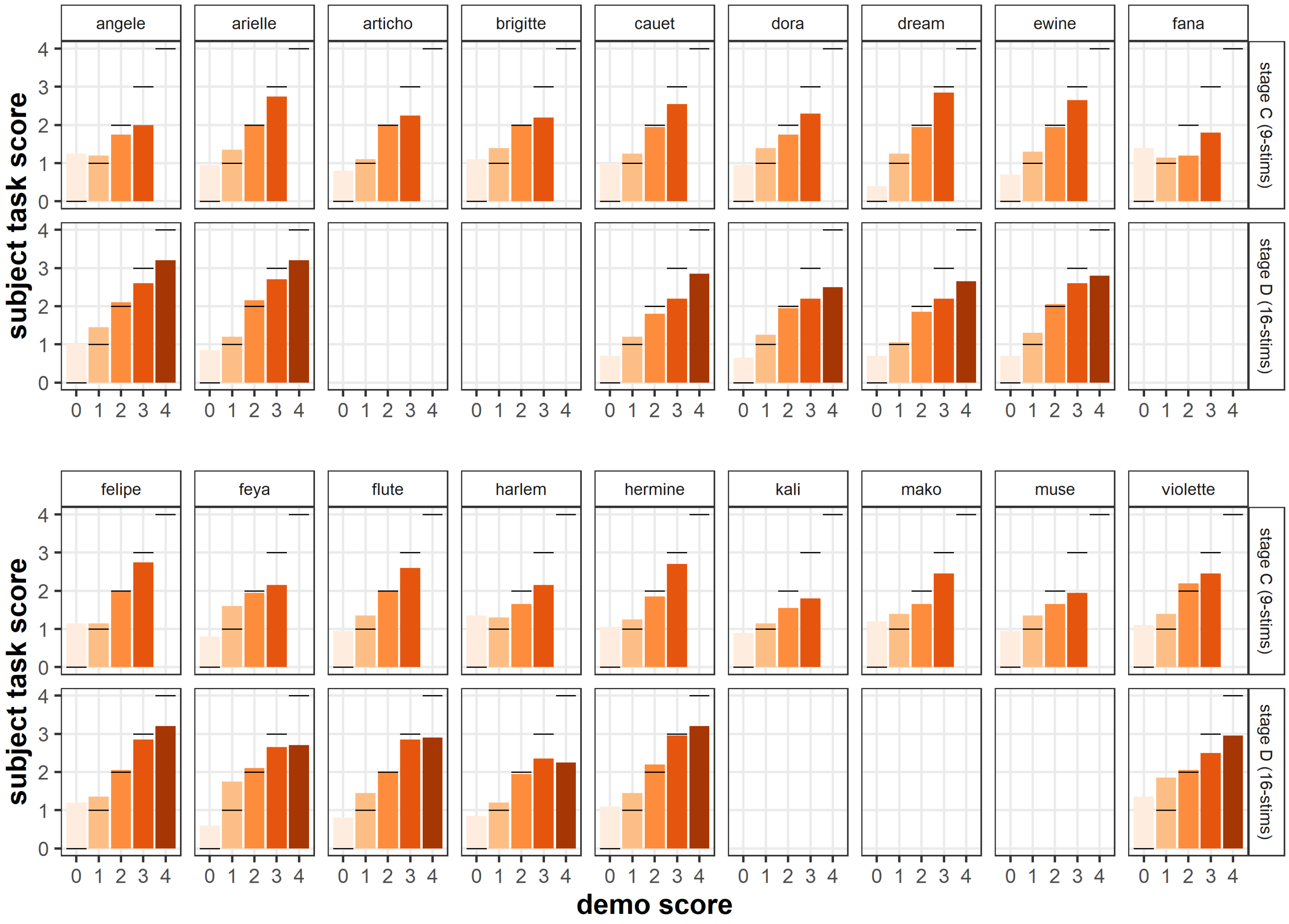


__Figure 5.6. Mean task score (number of targets found) by demonstration score, faceted by subject and stage.__ Horizontal black lines indicate demonstration score level.

Figure 5.7 displays the top classification level met by the group and by each subject on Stage C and Stage D. The population overall and many individual monkeys displayed the potential for cumulative culture in the transient cues condition of the current PFR Task. This was met numerically only; no subject (nor the group) was significantly above the threshold level on either stage. All classifications were at least at Level 2 indicating an overall responsiveness to the value of information provided.

As in Chapter 4, Stage D appears to be more sensitive for detecting potential for ratcheting; the group as a whole and several individuals displayed this ability on Stage D, whereas a single baboon only reached this classification level on Stage C.

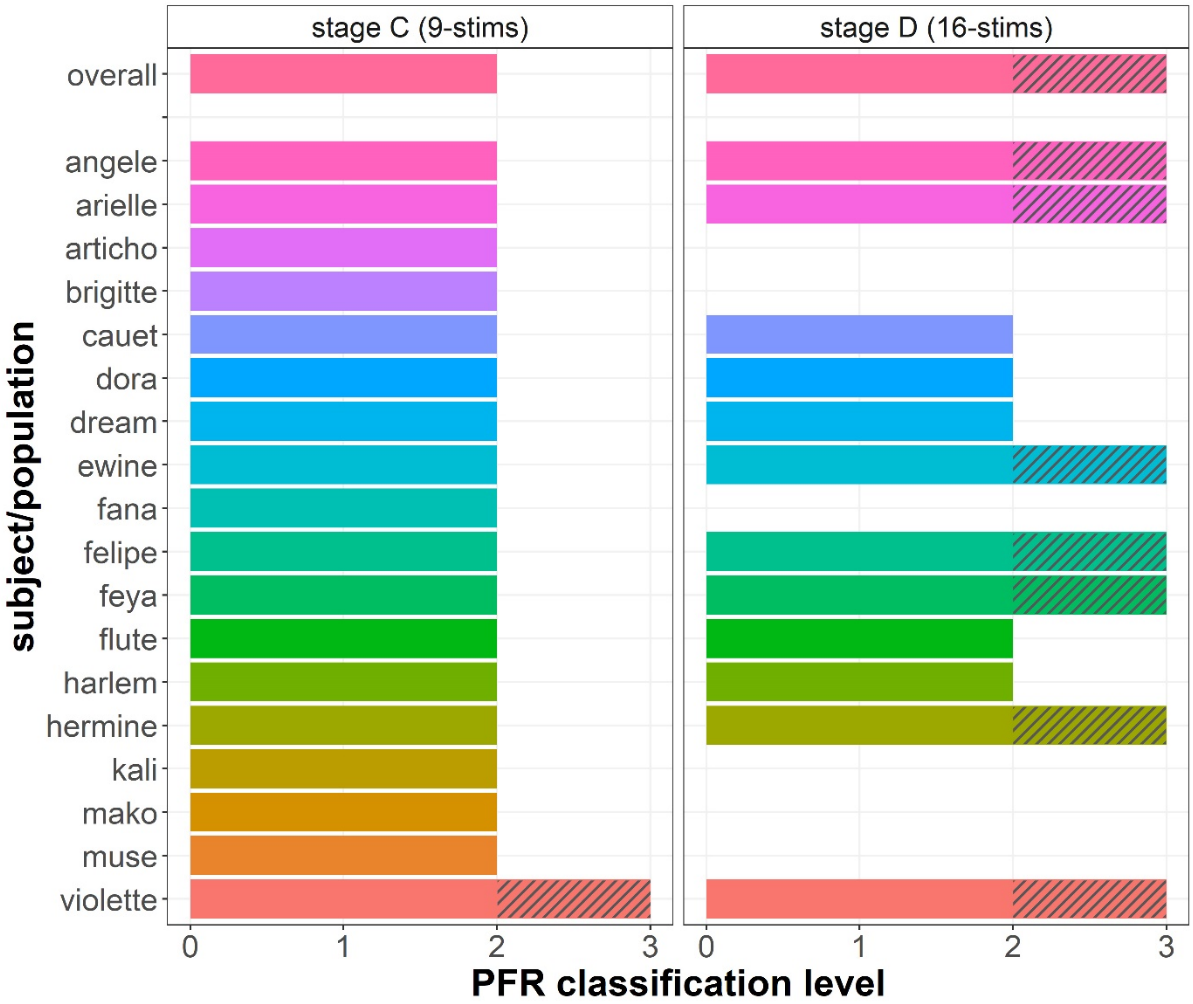


Figure 5.7. Top PFR Classification level met by each subject for both Stages C and D. *Shaded sections indicate that Level 3 was met numerically only (without significance).*

Finally, the results of the simulated transmission chains can be found in Figure 5.8 (group) and Figure 5.9 (individuals). Simulated performance displayed some increase in success over generations, with marked variation between subjects. This indicates the ability to accrue beneficial solutions to some degree. However the progress was very limited, with no performance plateaus approaching optimal (i.e., perfect application of the repeat rewarded and avoid unrewarded strategy). This is particularly clear on Stage D where progress falls proportionately shorter than optimum in relation to Stage C. Figure 5.10 displays the proportion of repetition of each information type separated by demonstration score. The use of information by the baboon population was clearly sub-optimal, and there appears to be a downward trend of repetition across demonstration score for both rewarded and unrewarded information types. This suggests that there was an increase in shift/explore responses as the demonstration value increased.

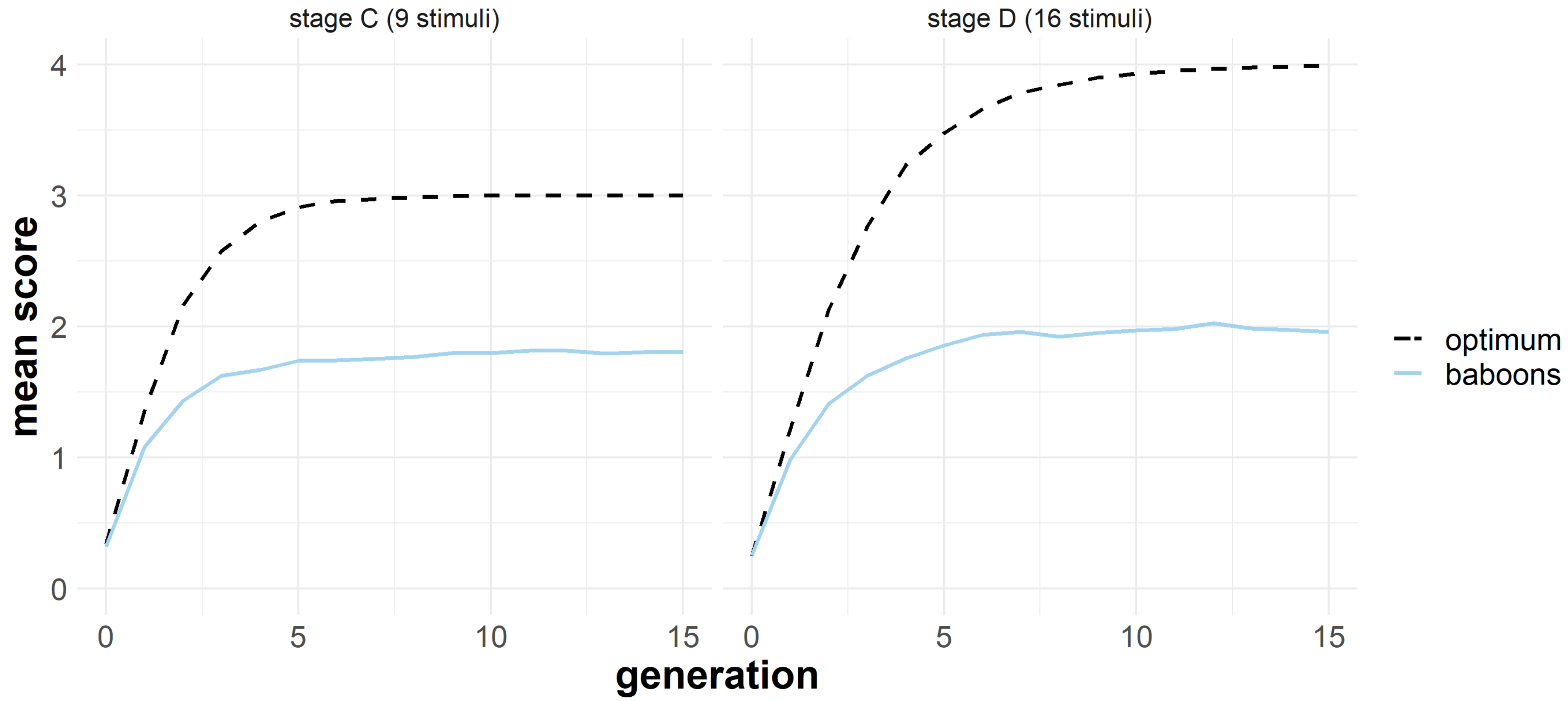


Figure 5.8. Simulated transmission chain performance (group level). *Simulations from both stages are displayed. The black dashed line represents optimal performance, i.e., fully consistent application of correct strategy. Chancel level for Stage C was 1/3 and for Stage D was 1/4.*

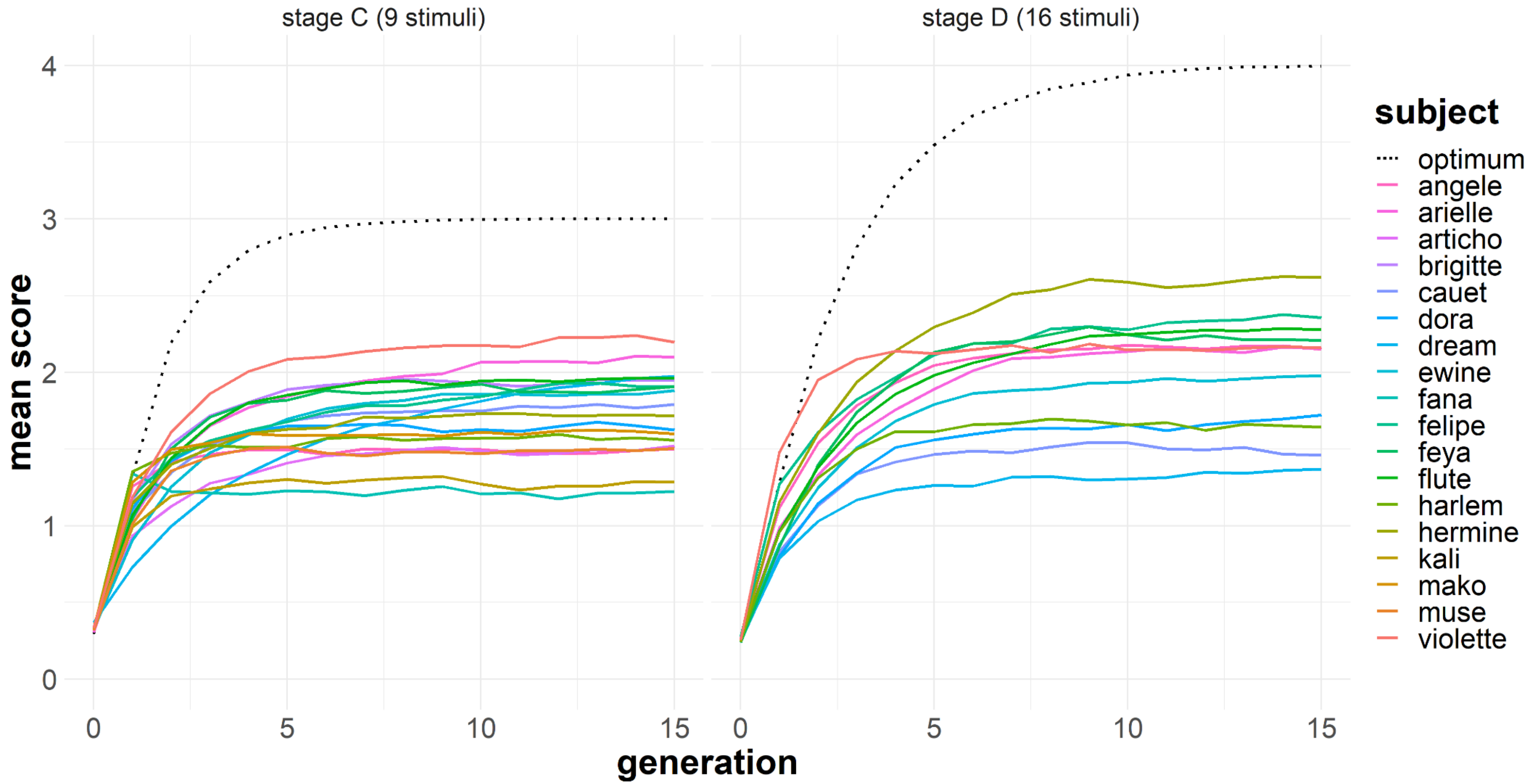


Figure 5.9. Simulated transmission chain performance for each subject. Simulations from both stages are displayed. *The black dotted line represents optimal performance, i.e., fully consistent application of correct strategy. Chancel level for Stage C was 1/3 and for Stage D was 1/4.*

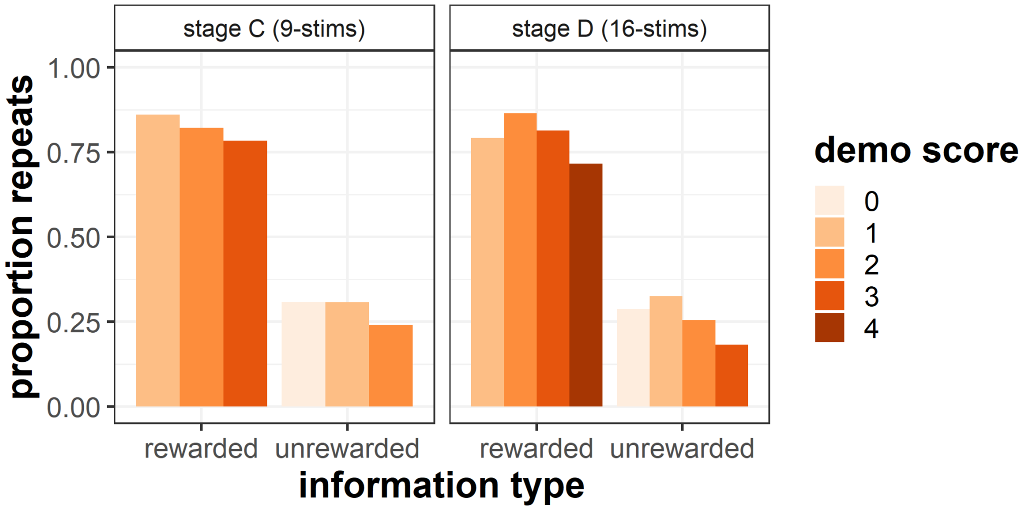


Figure 5.10. Proportion of repetition by information type, coloured by demonstration score and faceted by stage.

Discussion

Overall, the baboon population displayed the ability to accumulate beneficial information over learner generations at both the group and individual level. This provides evidence that a non-human primate species has some capacity for ratcheting even when working/short-term memory is somewhat taxed. The results display a clear ability to strategically use vicarious information to some degree, however this was also likely facilitated by the extensive relevant experience of the group (outlined in the Introduction).

However, the progress observed was extremely restricted, with no significant outperformance of the threshold associated with PFR Classification Level 3. Much like the capuchins in Chapter 4, although potential for ratcheting was displayed, it seems apparent that even the top performing monkeys met the criteria for ratcheting to a minimal degree, and accumulation would likely not progress beyond few episodes of transmission. The current study provides further evidence that it is higher scoring demonstrations (i.e., the scores that would be least likely by chance) that pose the greatest challenge for non-human primates on the PFR Task. This can be partially understood with reference to the increased likelihood of outperforming lower scoring demonstrations by chance compared to higher scoring ones (discussed in more detail in the Chapter 4 '*Discussion*').

The current set of results highlight the utility of requiring a flexible learning strategy in tests of capacities for cumulative culture. It was apparent from previous work that blanket repeat and avoid responses could be mastered separately (Claidière et al., 2014; Saldana et al., 2019); these studies displayed cultural evolutionary processes that continued over many generations of transmission, and which produced contrasting cultural lineages. However, the current set of results provide greater insight that both responses can also be deployed selectively by this population to generate cumulative improvements, i.e., the outperformance of the lower levels of demonstration found here. As such, once objective improvement was measured in the current study, rather than just copying fidelity and arbitrary changes, it has become clear that *cumulative* culture is probably heavily constrained, involving only very limited improvement over generations of transmission.

Unlike in the previous chapter, the burden on working/short-term memory may have also constrained performance in the current study. It is possible that providing visible cues may have engendered higher PFR classification levels or higher top scores when transmission chains were simulated in this experienced population. However, given the potential hindering effect from presenting the Kyoto capuchins with both visible and transient cues (discussed in Chapter 4), providing the current population with the cue condition that they had substantial experience with (Claidière et al., 2014; Saldana et al., 2019) was reasoned to provide the greatest potential for detecting ratcheting. A potential future direction could interleave the visible and transient cues conditions (the design we implemented with the children in Chapter 3) with a naïve non-human primate population, to test hypotheses more directly regarding the role of memory in generating a cumulative culture effect.

Overall, low precision information use again appears to have constrained the accumulation of success in a monkey species using the PFR Method. As accurate retention of rewarded information and consistent avoidance of unrewarded variants likely underlies ratcheting, inefficient use of vicarious information may at least partially explain the dearth of evidence for complex culture in non-human animals.

Chapter 6 General Discussion

Summary of Aims and Hypotheses

The primary goal of this thesis was to assess capacities for cumulative culture (or ‘ratcheting’) in non-human primates to gain insight into the distinctiveness of human culture. It ultimately aimed to present the first empirical evaluation of this capacity in two monkey species using the PFR (i.e., potential-for-ratcheting) Method outlined by Caldwell et al. (2020).

Although previous work has presented compelling candidates for cumulative culture in non-human animals (henceforth ‘animals’), the evidence remains either open to other explanations, or restricted to contexts that are tied to natural grouping tendencies (e.g., Jesmer et al., 2018; Sasaki & Biro, 2017). Nonetheless, this implies that the ability to accumulate beneficial modifications may be possible in very specific contexts in other species, and thus not entirely restricted to humans. Experimental work has further demonstrated the importance of context, finding that the cognitive capacities necessary for cumulative culture to emerge is dependent on task demands (Caldwell & Millen, 2009; Zwirner & Thornton, 2015). Together, this suggests that relatively simple forms of ratcheting may not require human-unique cognitive mechanisms and is more likely to emerge given conducive conditions.

Thus, basic forms of cumulative culture may be expected in a narrow range of contexts (involving favourable conditions) in animals, who may not have the same broad range of cognitive capacities as adult humans. However, whether animals have the potential to ratchet in contexts not tied to behavioural predispositions, remains unclear. As human culture appears unconstrained in the domains in which it can accumulate beneficial modifications, this is an important area of investigation for understanding human uniqueness.

A novel approach to test for cumulative culture ability, the PFR Method (Caldwell et al., 2020), allows clear assessment of this capacity in an abstract context, i.e., one that is not tied to species-specific abilities. Direct tests of the capacity for ratcheting in animals are scarce, and part of the reason for this may be that gathering data to sufficiently assess non-human cumulative culture is logistically challenging (discussed in Chapter 1). Using the PFR method, these issues are side-stepped by simulating inter-generational transmission at the individual level, reducing the organisational demands associated with typical generational replacement designs. The PFR Method has been shown to effectively elicit a cumulative culture effect in young children (Wilks et al., 2021), another population whose recruitment entails many of the same logistical issues.

As it is likely that cumulative culture is context-dependent, our main objective was to create optimal conditions which may provide the best opportunity for a ratchet effect to manifest in non-human primates. Learning conditions that are particularly conducive to ratcheting would likely involve a reduction or minimisation of any cognitive burdens associated with the task employed. Moreover, an *absence* of cumulative build-up across a range of testing circumstances that provided such facilitative conditions would potentially provide compelling negative evidence. Such a breadth of evidence would be required to more confidently support the assertion that the ability to engage in this process is *not* readily available to particular populations.

Overall, this thesis sought to assess whether optimal learning conditions may allow an unambiguous demonstration of cumulative culture (or elucidate the barriers to that) in non-human primates. Specifically, we utilised an efficient new approach to test whether animals have the potential to demonstrate this capacity under experimental conditions.

Key Findings, Assumptions and Limitations

The three major findings from this body of work will now be outlined; the key underlying methodological and theoretical assumptions will also be examined for each. As the methodologies employed were very similar throughout this thesis (i.e., all were presented on touchscreens and involved stimulus selection task designs) and all had closely related aims (i.e., identifying capacities for selective learning in non-human primates), the results will mainly be considered as a whole in this section. Limitations and potential alternative methodologies will also be described, where relevant.

Key Finding 1: Monkeys Have the Capacity to Selectively Learn in a Manner that May Theoretically Support Cumulative Culture.

The main focus of this thesis was whether a ratchet effect could be elicited in non-human primates using a task requiring selective use of information to accumulate success. Before applying the PFR Method to fully evaluate cumulative culture (Chapters 3-5), we performed an initial assessment of whether the same underlying selective strategy could be learned at a basic level (Chapter 2). A single strategy was therefore key to success on all the tasks employed in this thesis, involving repetition of rewarded and avoidance of unrewarded information.

Chapter 2 found that the tufted capuchin monkeys (*Sapajus apella*) tested appeared to struggle to grasp both contingencies of the selective strategy. An ability to learn and generalise the repeat-rewarded rule was displayed. However, concurrent avoidance of repeating unrewarded behaviour appeared difficult to master. This may have been the result of a pre-existing bias to repeat which favoured the response required in relation to rewarded information. Nonetheless, these results, together with the findings of Renner et al. (2021) with the same population, suggested some ratcheting based on this selective strategy may be possible in this species, given enough training and additional minimisation of cognitive demands.

Thus, in Chapters 4 and 5, we used a scaled-up version of this task and employed the PFR Method to directly test our main research question regarding capacities for cumulative culture in non-human primates. This involved multiple exposures per individual to task solutions of varying value, i.e., a range of levels of success, to mimic a range of task proficiencies, and thus, positions in a chain accumulating beneficial information. Responses were then used to infer the outcomes of within-individual linear transmission chains, providing an indication of the potential for ratcheting. In light of the difficulties found in mastering the basic task contingencies in Chapter 2, we first reduced pressure on memory in line with the aim to provide optimal conditions, discussed above. Subjects were first required to meet a performance criterion regarding accurate use of the optimal strategy before ratcheting was assessed.

Our results provided the first evidence that capuchin monkeys (theoretically) have the ability to engage in cumulative culture given favourable learning conditions; however, it is important to note that this required extensive training on the task contingencies, and not all individual and group level analyses indicated this level of ability. Furthermore, of those that displayed full potential for ratcheting, the pattern of results suggested that the progression would be very limited as the criterion for ratcheting was met only to a minimal degree. This implied that improvement would likely occur over only very few episodes of transmission, as when ratcheting was displayed, only lower value demonstrations were consistently improved upon; this will be discussed further below.

However, not all subjects displayed full potential; across the two groups of capuchins studied, we found a range of ability, from some subjects exhibiting poor information use (no ratcheting potential), others that derived some benefit from increasingly useful information (but no ratcheting up of solutions), and higher-performing individuals (suggesting full capacity for ratcheting). This range may be surprising as all individuals within these categories were presumably equally well-trained, as they had all passed the same performance criterion. Those that displayed the poorest performance, were found to gain no extra benefit from exposure to high value information compared to low. As such, these subjects showed the capacity to selectively use information to some degree by performing above chance, however no ratcheting potential was apparent. This will be returned to below.

Time restrictions prevented testing of one of our populations of capuchins in contexts of high memory load. However, despite extensive training being provided to the other group, none reached our performance criterion to be assessed for ratcheting in Chapter 4. This may have been due to the more cognitively challenging context (consistent with the findings of Wilks et al., 2021), or may have been attributable (either completely or at least in part) to order effects as a result of having learned to approach the task using methods that were, in the new context, detrimental to success (e.g., disregarding the information available from the animated “demonstration” in favour of simply selecting and avoiding stimuli on the basis of the enduring visual cue). However, Chapter 5 presented evidence that cumulative culture was indeed also possible in contexts involving some degree of memory pressure, in a population of well-trained Guinea baboons (*Papio papio*) who had prior experience of similar tasks involving transient cues. Thus it appears that contexts involving complete removal of memory load may not always be necessary to elicit a cumulative effect in animals. However, it is possible that this may only be most relevant to well-trained populations.

In spite of this difference in task presentation, the baboon population were found to perform similarly to the capuchins. Evidence for cumulative culture was found at the group level and for several individuals. However, again, we found that higher scoring demonstrations were not improved upon; this would likely generate a ratchet effect that would progress no further than a couple of transmission events. Also in line with the capuchins, many individuals did not display full capacity for cumulative culture, even though some sensitivity to the increasing value of information was found.

It should further be noted that some subjects failed to meet our performance criterion at all, which was necessary to allow ratcheting to be evaluated. This was despite considerable experience of task training trials, highlighting the range of inter-individual variation in ability to accumulate beneficial modifications in the species tested.

Although some definitions may require additional criteria to constitute ratcheting, such as requiring cumulative build-up of traits that are beyond the capacity of individual trial-and-error learning (e.g. Reindl et al., 2020), we suggest that the results of Chapters 4 and 5 demonstrably meet the minimum criteria set out by Mesoudi & Thornton (2018). However, the assumptions that this claim is based on are considered further below. Specifically, our findings indicate that the study species can selectively learn in a manner that supports ratcheting in the context of a stimulus selection task, in conditions of both low and high memory load.

However, low precision information use (i.e., inconsistently applying the optimal task strategy), particularly following higher value information, may have curtailed this potential. The higher value demonstrated information represented performances that are greater than might be expected by chance, and thus model outputs based on multiple episodes of cumulative transmission (i.e., the output of generations further down a ratcheting transmission chain). This suggests that cumulative culture in real life scenarios may have a very short history of improvement, due to difficulties in using information that itself developed as a result of a series of modifications.

Overall, these findings provide proof of concept that cumulative culture can be elicited using the PFR Method with non-human primates, using a task based on a simple form of selective information use. However, we would suggest that generalisation of these findings to natural conditions should be limited, as this effect was demonstrated in tightly controlled conditions and required substantial training.

Assumptions

To assess capacities for ratcheting, this thesis rested on the assumption that individuals must have the ability to strategically use information in a manner that supported discrete build-up of beneficial modifications. This supposition underlies the methodologies for each of the empirical chapters.

However, it is plausible that such strategic information use at the individual level is not necessary for all cases of cumulative culture. Instead, ratcheting could result from relatively simple success-biases regarding who to copy, where the behaviour of the most proficient individuals is preferentially replicated (Henrich, 2004). This could result in ratcheting if, for example, deviations away from perfect copying of the behaviour of high-success individuals (e.g., as a result of copying error) created variations of the trait at the population level; some individuals may randomly become more, and some might become less successful than the original high-success model. Thus, improvements can accumulate (without requiring strategic information use, just copying), with the associated individuals becoming the preferred models for the next generation of success-biased learners. Overall, this can gradually draw the functionality of the trait upwards at the population level due to copying deviations coupled with model selection bias. By this theory, even if most case of transmission resulted in a loss of proficiency, the success-bias would operate to offset this by directing copying toward the few high success individuals.

Although the process described above would not depend on selective use of information generating outperformance of demonstrations through strategic copying (as required to demonstrate ratcheting in the current thesis), it would nonetheless require selectivity in choosing the best model to copy from a range of candidates. The ability to discriminate between demonstrations of varying levels of success, and to direct one's copying accordingly, arguably involves cognitive challenges even more significant than those involved in selective information use; especially given that exposures to competing demonstrations are likely to be separated in time and space (e.g., see Wilks et al., under revision).

Furthermore, although the improvements in performance arising as a result of success-bias emerge as a group level phenomenon, they are also logically underpinned by a pattern of performance at the individual level which is effectively evaluated in the current thesis: that being the need to derive greater benefit following exposure to higher performing models (even if not requiring outperformance of those models). This would be necessary to benefit from the success-bias, otherwise copying more successful individuals would provide no additional benefit compared to selecting a model at random. This is captured by our PFR Classification Level 2 (Table 3.2) which requires performance to be positively related to demonstration score, but does not necessarily require improved performance as a result of high precision selective copying (which is captured by our Level 3). Therefore, the PFR Method can also provide some indication of capacity for ratcheting according to this theory.

This work has also rested on the assumption that non-human primates are a prime candidate for generating a ratchet effect. This is due to their extensive cultural repertoire (e.g., Whiten et al., 1999), and their phylogenetic proximity to humans which may suggest some fundamental shared capacity. However, as some of the most convincing evidence comes from homing pigeons (*Columba livia*) that group in order to travel long distances (Sasaki & Biro, 2017), such species could have been more fruitful study subjects due to easier access to large numbers, and therefore traditional transmission chain studies could potentially have been run. The benefits of this are discussed further in relation to the second key finding below. However, as the evidence from such species is not suggestive of the capacity to recognise and benefit from vicarious information, but more likely an incidental outcome of grouping behaviour (Jesmer et al., 2018; Sasaki & Biro, 2017), there is little reason to suggest that these species are particularly good candidates for displaying a capacity for cumulative culture using a more abstract task. Furthermore, the success of the current body of work displays the effectiveness of the PFR Method for testing populations that have received much attention in the literature surrounding the phenomenon of cumulative culture, non-human primates, but very little direct testing of their ratcheting capabilities.

Key Finding 2: The PFR Method is an Effective Means of Assessing Potential for Cumulative Culture.

Another major finding from the current project is that the PFR Method is suitable for use with non-human primates. This approach uses individual level responses to information of a range of values, to infer the outcomes of linear transmission chains. The task itself can vary, provided it allows objective increases in success to be identified and no carry-over effects (Caldwell et al., 2020). Chapter 3 first provided validation of a new touchscreen-based task for use with the PFR Method, through replicating the results of an established, non-electronic based task that applied this method with a population of young children (Wilks et al., 2021). The novel instantiation of the PFR Method described in the current thesis is the first to allow implementation with both humans and any other species that can be trained to use touchscreen computers.

Following this validation, the touchscreen-based task was then successfully employed with non-human primates in Chapters 4 and 5 allowing the potential for cumulative culture to be assessed, as described above, and providing the first validation of the PFR Method with a non-human species. This adds a third application of this methodological approach, following Wilks et al. (2021) and Wilks et al. (under revision), to the literature, and the first with a non-human.

Our overarching aim to assess capacities for cumulative culture further benefitted from the utility of the PFR Method as a nuanced evaluation of this ability was made possible; for example, even given some evidence of selective vicarious learning in an individual, it was found that this does not necessarily equate to the ability to engage in cumulative culture (Chapters 3-5, described above). This distinction was identified due to the PFR Method's sensitivity to differences between the general capability for strategic social learning, and the extension of this capability allowing ratcheting of this information up by accumulating beneficial modifications.

Assumptions

In employing the PFR Method, I have taken the position that we can extrapolate from the behaviour of individuals on an abstract task to infer the outcomes of multi-individual transmission chains (Chapters 3-5).

Firstly, there are many benefits to this approach. The potential for ratcheting of two monkey species was evaluated in the current project, however any species that can be trained to use a touchscreen could theoretically be tested on this paradigm, as noted above. As the task is abstract and relies on general (i.e., not species specific) learning abilities, and requires only simple motor responses, relatively fair comparisons across species are possible using this method. Moreover, versions of the PFR Task employed here, but using physical objects, may allow inclusion of non-touchscreen capable species (see below for an example of how this could be operationalised). Evaluating a wide range of taxa could provide insights into the spread of the capacity to strategically use information to support cumulative culture. Much comparative cognitive research focusses on non-human primates, however, many other species have displayed evidence of culture, and problem-solving abilities (e.g., Hunt & Gray, 2003; Rendell & Whitehead, 2001) and therefore may also be good candidates for displaying the capacity for ratcheting, particularly given optimal conditions. Evidence of the potential for cumulative cultural evolution across a wider range of species would suggest that this phenomenon is not specific to the primate order but has a wider taxonomic range, generating a powerful insight into the seeming uniqueness of human cultures.

Additionally, previously there was no efficient method for nuanced differentiation between the ability to selectively learn, and the capacity to ratchet, much less one that could be presented across a range of non-human species. Experiments explicitly testing capacities for cumulative culture in future, theoretically or otherwise, now have a method to evaluate whether selective learning from vicarious information would ultimately result in cumulative improvements, determining whether ratcheting capacity can be soundly inferred. The PFR Method provides an effective means for doing this rather than simply extrapolating from a basic ability to strategically use information, such as found in social learning experiments using win-stay, lose-shift tasks (e.g., Chapter 2; Monfardini et al., 2012, 2014). This is illustrated by our PFR Classification System (Table 3.2), which represents a continuum of potential for ratcheting, distinguishing between levels of increasingly sophisticated selective information use. Thus, capacities to obtain some benefit from vicarious information by performing above chance (Level 1), can be differentiated from deriving additional benefits from increasingly beneficial information (Level 2), which can in turn be differentiated from using that information to generate improvements (Level 3). Thus, our results highlight the utility of the PFR Method for elucidating the degree of potential for cumulative culture.

However, there are also limitations to the PFR Method approach. For example, our studies sacrificed ecological validity in favour of greater experimental control. This approach was taken to increase the feasibility of assessing ratcheting in non-human populations, and to create a task that was not tied to natural predispositions. However, although this may be more difficult to implement, it would be possible to modify the PFR Task used in Chapters 3-5 to create learning conditions that are more relevant to problems that could be encountered in the wild. For example, a 3D ‘artificial foraging patch’ could be created, involving a board displaying a matrix of small opaque containers, a proportion of which contained real food rewards, in line with the stages of the PFR Task used here (Table 3.1). The containers represent the stimuli, and could either provide transient cues, e.g., they could be closed after selection during the demonstration, or visible cues, e.g., they could remain slightly open, revealing part of their contents during the test trials (much like the treasure chest stimuli used by Wilks et al., 2021). An experimenter would be required to bait the required number of containers for each individual trial and some containers would be opened, either by the experimenter or a conspecific, to provide a demonstration. It would be critical to ensure distinctive arrays of stimuli were presented however, to ensure a clear understanding of the independence of the reward distributions between successive problems (established in our touchscreen task through extensive training). The subject would then be free to indicate the containers they would like to search following this demonstration. This direct association between stimuli and reward may engender greater motivation, attention or understanding of the contingencies of the task (Wood & Whiten, 2017), potentially generating a higher inferred propensity for ratcheting. This would be an interesting avenue for future research, however, was not adopted here due to the significantly greater organisational demands. Moreover, due to the automated system used to conduct experiments with the baboons tested in Chapter 5, this would not have been feasible to run with this population. Additionally, it is possible that this approach could be *more* confusing than the simple touchscreen presentation utilised here. Less abstraction, and reliance on intuitive understanding rather than training, might also introduce task demands that could not be considered equivalently challenging across a range of species.

An additional limitation is that inferring capacities for ratcheting from single participants, or small numbers of individuals, may have lower external validity compared to transmission chains where each generation involves a new individual. "Real" social transmission could of course have been more closely approximated by assembling chains of individual learners or could have been studied directly through longitudinal observations of animals in natural conditions. So, for example, the construction of traditional transmission chains (involving different participants at each generation) could also have been attempted with non-human primates instead of drawing inferences from individual level responses (Chapters 4-5). The replacement of each generation with a new, naïve individual would also allow a greater range of tasks to be employed; the PFR Method necessarily requires a task involving no carry-over effects which would typically necessitate a substantial degree of abstraction, reducing the range of possible task options. However, observational evidence of cumulative culture from long-term studies of the natural behaviour of non-human primates would of course provide the strongest evidence that this phenomenon is not restricted to humans, not just in theory but in practice.

However, for our purposes, we considered the PFR Method a more attractive approach than either of these alternatives. The benefits to this method, outlined in Chapter 1 (section '*The PFR Method*'), justified its use due to the substantial reduction in organisational demands, relative to traditional transmission chain designs; see Horner et al. (2006), described in Chapter 1, for an illustration of the challenges in controlling the construction of chains with real generational turnover with chimpanzees (*Pan troglodytes*).

Secondly, and fairly obviously, long-term observations that have the scope for capturing cumulative cultural evolution in natural conditions were out of the reach of this PhD project due to time constraints. As long-term field sites have studied natural populations for decades, this is clearly a time-intensive endeavour, especially since candidate behaviours for ratcheting, even those which have been studied over many years, invariably remain open to interpretation (Dean et al., 2014; Tennie et al., 2009).

Finally, I have presented vicarious information as a suitable proxy for social information in Chapters 2-5; this will be examined in relation to the third key finding below.

Overall, despite the limitations mentioned here, I believe the PFR Method approach is a valuable test of cumulative culture over and above the alternatives, due to mitigation of common barriers to assessing non-human potential for cumulative culture in this field of research.

Key Finding 3: Information is Learned and Used Equally Well from Social Sources and Individual Exploration.

Our test of the foundational ability to engage in the repeat-rewarded, avoid-unrewarded strategy (Chapter 2, discussed above) concurrently investigated whether the learning and use of this strategy differed depending on whether it was derived from social or individual sources. This was operationalised using a between-groups design, where the experimenter performed the information trial/demonstration in the social learning condition, and the subjects themselves in the individual learning condition. This was followed by test trials where the subject's use of this information was monitored. As we have posited that strategic information use is fundamental to cumulative cultural evolution, directly comparing these learning conditions provided valuable insights, particularly due to the emphasis placed on the difficulties of animals in learning socially in the cumulative culture literature (e.g., Boyd & Richerson, 1996; Tennie et al., 2009). Moreover, adequate comparisons of information use from social and individual sources are rare. Thus, the information between these source conditions was strictly matched in value in Chapter 2, allowing a rigorous comparison regarding the influence of source on the use of learner information.

Little difference was found in the learning and use of the selective strategy between individuals assigned to either social or individual learning conditions, suggesting that capuchins experience no underlying difficulty in utilising information not generated by their own activity. This remained true when presented with problems with a novel structure which required the same strategy (see Chapter 2, '*Transfer*' sections). Overall, these results implied that information source is not a differentiating factor in the context of the stimulus selection tasks used in the current thesis (although insufficient power remains a potential explanation for this null result). This indicated that, for the purpose of these tasks, the source of information was largely inconsequential, and therefore not a key consideration when designing the tasks used Chapters 3-5. More broadly, it also demonstrated that use of social information is not a fundamental barrier to flexible information use in monkeys. As mentioned above, it is widely assumed that cultural differences between humans and non-human primates are a consequence of differential social learning capacities. However when the information available between social and individual learning conditions was strictly matched, we found that this was unsubstantiated, instead suggesting that general information use may inhibit strategic information use at a basic level.

Assumptions

The context in which the information is provided in the social condition of the win-stay, lose-shift task of Chapter 2 (i.e., the experimenter simply touching one of the stimuli on the touchscreen), assumes that this represents a "social" demonstration that sufficiently taps into the essential differences between social and individual learning. However, due to the simplistic nature of the task, it could be argued that this is an unrealistic representation of a social learning context in real life scenarios. It is possible that providing information in a richer social environment may have influenced motivation or attention, and thus elicited greater variation between these source conditions.

However, we believe this was a necessary trade-off in order to hold informational value constant across the conditions. We see this as a strength of our method, developing our understanding of barriers to strategic information use in non-human primates, as previous research into the social learning of animals has arguably failed to suitably control for this (e.g., Nagell et al., 1993). However, we accept that generalisation of these results to the real world may be limited and would welcome comparisons between social and individual learning conditions involving greater ecological validity, whilst stressing the importance of providing matched information between conditions. We also acknowledge that variation between information sources may become more apparent in different contexts, such as those that require an understanding of the actions of a social model in order to infer the function of a cultural product.

This assumption is also relevant to the vicarious presentation of information in the touchscreen-based PFR Task, responses to which were used to infer the outcomes of transmission chains in Chapters 3-5, mentioned above; this could also be argued to constitute a less convincing 'social' scenario. The assumption here is that the presentation of vicarious information (using an animated demonstration in this case) represented an appropriate means of operationalising information acquired as the result of another individual's activity. Moreover, in the visible cues condition presented to the capuchins in Chapter 4, it was possible for the subjects to pay little attention to the animated demonstration, but instead to rely on the enduring cues that remained visible during the test trials. However, unlike Chapter 2, we were not primarily interested in the influence of source in this context, but rather, general abilities to engage in a flexible strategy to use information that had been acquired from a source other than one’s own activity. Moreover, Chapter 2 revealed that, although there were barriers to task success, these were not source-specific, but applied equally across the two source conditions. As such, our focus for Chapters 3-5 was on the ability to selectively learn from information not generated by the individual themselves, regardless of where that information had originated. It is perhaps worth noting that numerous experiments involving social learning with humans, including many investigations of cumulative culture, do not involve the physical presence of a social model (e.g., Derex et al., 2012; Jiménez & Mesoudi, 2020; Kirby et al., 2008). Furthermore, many examples of social learning in natural conditions in the wild mentioned in Chapter 1, e.g., following pheromone trails in ants (Czaczkes et al., 2015) and foraging in milk bottles in birds (Fisher & Hinde, 1949), also involve similarly vicarious information, without necessarily requiring contact with the individual responsible for generating the information. Therefore, since the research presented in Chapters 3-5 was never intended to represent an investigation of social learning mechanisms per se, the vicarious presentation of information would seem to represent a perfectly valid means of studying the broad capacity to learn and use information gained from external sources.

Future Directions and Outstanding Questions

Chapters 4 and 5 provide proof of principle that non-human primates have the capacity for cumulative culture, at least under very specific conditions. Further to those already discussed, here I will outline potential avenues for future work that this project has stimulated, and outstanding questions.

Regarding context, the current study successfully compiled data from different populations using conditions of varying memory load (visible vs. transient cue conditions), however we had originally aspired to compare ratcheting capacity across these conditions using a within-groups design (Chapter 4). This would have constituted a more powerful approach for establishing the influence of memory load on ratcheting, and might even have shed some light on a leading theory regarding the distinctiveness of human culture (see Balter, 2010). However, practical considerations, including time constraints, limited our ability to investigate this systematically. Clear contrasts in ratcheting potential between conditions are a particularly useful way to establish negative evidence (Caldwell et al., 2020), e.g., finding that one group cannot ratchet when memory load is high, but can when this burden is removed, would robustly suggest that high memory load is a limiting factor. Future research may benefit from interleaving these conditions within experimental sessions, as opposed to sequential exposure. This may reduce the potential for interference effects caused by confounding condition and order of presentation (as may have occurred with the Kyoto capuchins in Chapter 4). We believe this is an important area of research in line with the emphasis placed on context in the current body of work. Such direct comparisons would represent an effective means by which to ascertain the role of a particular context or task demand in the emergence of cumulative culture.

However, as evidence of ratcheting was demonstrated by the baboons in Chapter 5, the degree of memory load in the current transient cues condition clearly was not a fundamental barrier to ratcheting (although this potentially depends on substantial prior experience). Determining the extent of pressure on working memory that may yet allow the capacity for cumulative culture in well-trained populations may further our understanding of both memory capacity more generally, and the upper capacities for cumulative culture in animals. This would be particularly relevant to theories positing a key role for memory in developing distinctively human culture (e.g., Wynn & Coolidge, 2007).

Variations to the conditions for learning would further allow the role of context and task demands in eliciting a cumulative effect to be understood, as demonstrated in relation to memory load in Chapter 3 and Wilks et al. (2021). For example, contexts requiring greater control of attention may be expected to reduce the potential for ratcheting in certain populations, such as non-human primates, who may have lower control over this capacity relative to humans (Beran et al., 2016; Carruthers, 2013). This could be manipulated using a variation of the current PFR Task involving two conditions requiring high and low attentional control. High control contexts may involve some form of distraction (e.g., a small object/shape moving from the top left to the top right of the touchscreen) during test trials, whereby the subject must resist interference in order to respond optimally. This could be conducted under the visible cues condition to ensure other sources of cognitive load, i.e., memory, are minimised. A relatively short time limit for responses to be given would also ensure that active attention was required to be successful. This condition could be interleaved with trials with no such interference, and the potential for ratcheting across the two contexts compared. Whether non-human primates may be capable of displaying ratcheting in varying circumstances of cognitive load may develop our understanding of human-unique aspects of cognition that affect cultural accumulation.

Although it was not a focus here, direct comparisons between children and non-human primates would be enlightening to this area of research; this is a particular strong point of the PFR Method (Caldwell et al., 2020). As age related developments may increase the range of contexts in which children demonstrate potential for ratcheting (Wilks et al., 2021), it would be valuable to investigate at what age, and in which contexts, departures from the performances of animals are observed in children.

Finally, evaluation of cumulative culture using the PFR Method paradigm with nonhumans (although not children, as shown in Chapter 3) requires substantial training. However, whether ratcheting may be possible from the spontaneous behaviour of animals (such as found in many generational replacement studies with humans) remains an open question.

Concluding Remarks

This thesis supports the theory that cumulative culture is not precluded in non-human species but probably occurs only in very specific contexts and is likely to involve only relatively limited development of beneficial modifications. As such, cumulative culture likely hinges on the interaction between cognitive capacities and conditions for learning. This potentially provides an explanation for the rarity of this phenomenon in other animals. In contrast, human-unique cognitive capacities (likely not specifically confined to the social domain) may allow ratcheting across a much broader range of contexts.

Finally, the PFR Method (Caldwell et al., 2020) provides a sound and relatively efficient means to assess non-human capacity for cumulative culture. This approach provides the potential for many fruitful avenues of future research.

References

Acerbi, A., Tennie, C., & Nunn, C. L. (2011). Modeling imitation and emulation in constrained search spaces. *Learning and Behavior*, *39*(2), 104–114. https://doi.org/10.3758/s13420-010-0009-z

Akins, C. K., & Zentall, T. R. (1996). Imitative learning in male Japanese quail (*Coturnix japonica*) using the two-action method. *Journal of Comparative Psychology*, *110*(3), 316–320. https://doi.org/10.1037/0735-7036.110.3.316

Alem, S., Perry, C. J., Zhu, X., Loukola, O. J., Ingraham, T., Sovik, E., & Chittka, L. (2016). Associative Mechanisms Allow for Social Learning and Cultural Transmission of String Pulling in an Insect. *PLoS Biology*, *14*(10), e1002564. https://doi.org/10.1371/journal.pbio.1002564

Allen, J. A. (2019). Community through Culture: From Insects to Whales. *BioEssays*, *41*(11), 1900060. https://doi.org/10.1002/bies.201900060

Aplin, L. M. (2019). Culture and cultural evolution in birds: a review of the evidence. In *Animal Behaviour* (Vol. 147, pp. 179–187). Academic Press. https://doi.org/10.1016/j.anbehav.2018.05.001

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/https://doi.org/10.1037/xge0000959

Balter, M. (2010). Did working memory spark creative culture? *Science*, *328*(5975), 160–163. https://doi.org/10.1126/science.328.5975.160

Bandini, E., & Harrison, R. A. (2020). Innovation in chimpanzees. *Biological Reviews*, *95*(5), 1167–1197. https://doi.org/10.1111/brv.12604

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328. https://doi.org/10.3389/fpsyg.2013.00328

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting Linear Mixed-Effects Models using lme4*. http://arxiv.org/abs/1406.5823

Battesti, M., Moreno, C., Joly, D., & Mery, F. (2015). Biased social transmission in Drosophila oviposition choice. *Behavioral Ecology and Sociobiology*, *69*(1), 83–87. https://doi.org/10.1007/s00265-014-1820-x

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature*, *456*(7219), 245–249. https://doi.org/10.1038/nature07538

Bejan, A., Charles, J. D., & Lorente, S. (2014). The evolution of airplanes. *Journal of Applied Physics*, *116*(4), 044901. https://doi.org/10.1063/1.4886855

Bender, C. E., Herzing, D. L., & Bjorklund, D. F. (2009). Evidence of teaching in atlantic spotted dolphins (*Stenella frontalis*) by mother dolphins foraging in the presence of their calves. *Animal Cognition*, *12*(1), 43–53. https://doi.org/10.1007/s10071-008-0169-9

Beran, M. J., Menzel, C. R., Parrish, A. E., Perdue, B. M., Sayers, K., Smith, J. D., & Washburn, D. A. (2016). Primate cognition: attention, episodic memory, prospective memory, self-control, and metacognition as examples of cognitive control in nonhuman primates. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(5), 294–316. https://doi.org/10.1002/wcs.1397

Beran, M. J., Perdue, B. M., Futch, S. E., Smith, J. D., Evans, T. A., & Parrish, A. E. (2015). Go when you know: Chimpanzees’ confidence movements reflect their responses in a computerized memory task. *Cognition*, *142*, 236–246. https://doi.org/10.1016/j.cognition.2015.05.023

Bernstein-Kurtycz, L. M., Hopper, L. M., Ross, S. R., & Tennie, C. (2020). Zoo-Housed Chimpanzees Can Spontaneously Use Tool Sets But Perseverate on Previously Successful Tool-Use Methods. *Animal Behavior and Cognition*, *7*(3), 288–309. https://doi.org/10.26451/abc.07.03.03.2020

Bird, C. D., & Emery, N. J. (2009). Insightful problem solving and creative tool modification by captive nontool-using rooks. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(25), 10370–10375. https://doi.org/10.1073/pnas.0901008106

Blanco, N. J., & Sloutsky, V. M. (2020). Attentional mechanisms drive systematic exploration in young children. *Cognition*, *202*, 104327. https://doi.org/10.1016/j.cognition.2020.104327

Boesch, C., Head, J., & Robbins, M. M. (2009). Complex tool sets for honey extraction among chimpanzees in Loango National Park, Gabon. *Journal of Human Evolution*, *56*(6), 560–569. https://doi.org/10.1016/j.jhevol.2009.04.001

Boesch, C., Kalan, A. K., Mundry, R., Arandjelovic, M., Pika, S., Dieguez, P., Ayimisin, E. A., Barciela, A., Coupland, C., Egbe, V. E., Eno-Nku, M., Michael Fay, J., Fine, D., Adriana Hernandez-Aguilar, R., Hermans, V., Kadam, P., Kambi, M., Llana, M., Maretti, G., … Kühl, H. S. (2020). Chimpanzee ethnography reveals unexpected cultural diversity. *Nature Human Behaviour*, *4*(9), 910–916. https://doi.org/10.1038/s41562-020-0890-1

Boinski, S., Quatrone, R. P., & Swartz, H. (2000). Substrate and Tool Use by Brown Capuchins in

Suriname: Ecological Contexts and Cognitive Bases. In *American Anthropologist* (Vol. 102, pp. 741–761). WileyAmerican Anthropological Association. https://doi.org/10.2307/684197

Bonini, L., & Ferrari, P. F. (2011). Evolution of mirror systems: A simple mechanism for complex cognitive functions. *Annals of the New York Academy of Sciences*, *1225*(1), 166–175. https://doi.org/10.1111/j.1749-6632.2011.06002.x

Bonnie, K. E., & de Waal, F. B. M. (2007). Copying without rewards: socially influenced foraging decisions among brown capuchin monkeys. *Animal Cognition*, *10*(3), 283–292. https://doi.org/10.1007/s10071-006-0069-9

Bonnie, K. E., Milstein, M. S., Calcutt, S. E., Ross, S. R., Wagner, K. E., & Lonsdorf, E. V. (2012). Flexibility and Persistence of Chimpanzee (*Pan Troglodytes*) Foraging Behavior in a Captive Environment. *American Journal of Primatology*, *74*(7), 661–668. https://doi.org/10.1002/ajp.22020

Bono, A. E. J., Whiten, A., van Schaik, C., Krützen, M., Eichenberger, F., Schnider, A., & van de Waal, E. (2018). Payoff- and Sex-Biased Social Learning Interact in a Wild Primate Population. *Current Biology*, *28*(17), 2800-2805.e4. https://doi.org/10.1016/j.cub.2018.06.015

Boyd, R., & Richerson, P. J. (1996). Why culture is common, but cultural evolution is rare. *Proceedings of the British Academy*, *88*, 77–93. https://doi.org/citeulike-article-id:1339814

Brewer, S. M., & McGrew, W. C. (1990). Chimpanzee Use of a Tool-Set to Get Honey. *Folia Primatologica*, *54*(1–2), 100–104. https://doi.org/10.1159/000156429

Brown, C., & Laland, K. N. (2003). Social learning in fishes: a review. *Fish and Fisheries*, *4*(3), 280–288. https://doi.org/10.1046/j.1467-2979.2003.00122.x

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science*, *307*(5712), 1118–1121. https://doi.org/10.1126/science.1105783

Buskell, A. (2020). Cumulative culture and complex cultural traditions. *Mind & Language*. https://doi.org/10.1111/mila.12335

Caldwell, C. A. (2018). Using Experimental Research Designs to Explore the Scope of Cumulative Culture in Humans and Other Animals. *Topics in Cognitive Science*, *12*(2), 673–689. https://doi.org/10.1111/tops.12391

Caldwell, C. A., Atkinson, M., Blakey, K. H., Dunstone, J., Kean, D., Mackintosh, G., Renner, E., & Wilks, C. E. H. (2020). Experimental assessment of capacities for cumulative culture: Review and evaluation of methods. *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(1), e1516. https://doi.org/10.1002/wcs.1516

Caldwell, C. A., Atkinson, M., & Renner, E. (2016). Experimental approaches to studying cumulative cultural evolution. *Current Directions in Psychological Science*, *25*(3), 191–195. https://doi.org/10.1177/0963721416641049

Caldwell, C. A., Cornish, H., & Kandler, A. (2016). Identifying innovation in laboratory studies of cultural evolution: rates of retention and measures of adaptation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *371*(1690), 20150193. https://doi.org/10.1098/rstb.2015.0193

Caldwell, C. A., & Millen, A. E. (2008). Experimental models for testing hypotheses about cumulative cultural evolution. *Evolution and Human Behavior*, *29*(3), 165–171. https://doi.org/10.1016/j.evolhumbehav.2007.12.001

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

Caldwell, C. A., Renner, E., & Atkinson, M. (2018). Human Teaching and Cumulative Cultural Evolution. *Review of Philosophy and Psychology*, 1–20. https://doi.org/10.1007/s13164-017-0346-3

Call, J., Carpenter, M., & Tomasello, M. (2005). Copying results and copying actions in the process of social learning: chimpanzees (*Pan troglodytes*) and human children (*Homo sapiens*). *Animal Cognition*, *8*(3), 151–163. https://doi.org/10.1007/s10071-004-0237-8

Carruthers, P. (2013). Evolution of working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(SUPPL2), 10371–10378. https://doi.org/10.1073/pnas.1301195110

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: Biological Sciences*, *281*(1797), 20141541. https://doi.org/10.1098/rspb.2014.1541

Clay, Z., & Tennie, C. (2018). Is Overimitation a Uniquely Human Phenomenon? Insights From Human Children as Compared to Bonobos. *Child Development*, *89*(5), 1535–1544. https://doi.org/10.1111/cdev.12857

Coolidge, F. L., & Wynn, T. (2001). Executive functions of the frontal lobes and the evolutionary ascendancy of *Homo sapiens*. *Cambridge Archaeological Journal*, *11*(2), 255–260. https://doi.org/10.1017/S0959774301000142

Cosmides, L., & Tooby, J. (2015). Adaptations for Reasoning About Social Exchange. In *The Handbook of Evolutionary Psychology* (pp. 1–44). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119125563.evpsych225

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

Custance, D., Whiten, A., & Fredman, T. (1999). Social learning of an artificial fruit task in capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology*, *113*(1), 13–23. https://doi.org/10.1037/0735-7036.113.1.13

Czaczkes, T. J., Grüter, C., & Ratnieks, F. L. W. (2015). Trail Pheromones: An Integrative View of Their Role in Social Insect Colony Organization. *Annual Review of Entomology*, *60*(1), 581–599. https://doi.org/10.1146/annurev-ento-010814-020627

Danchin, E., Nöbel, S., Pocheville, A., Dagaeff, A. C., Demay, L., Alphand, M., Ranty-Roby, S., Van Renssen, L., Monier, M., Gazagne, E., Allain, M., & Isabel, G. (2018). Cultural flies: Conformist social learning in fruitflies predicts long-lasting mate-choice traditions. *Science*, *362*(6418), 1025–1030. https://doi.org/10.1126/science.aat1590

Davis, S. J., Vale, G. L., Schapiro, S. J., Lambeth, S. P., & Whiten, A. (2016). Foundations of cumulative culture in apes: improved foraging efficiency through relinquishing and combining witnessed behaviours in chimpanzees (*Pan troglodytes*). *Scientific Reports*, *6*(1), 35953. https://doi.org/10.1038/srep35953

Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B., & Laland, K. N. (2012). Identification of the social and cognitive processes underlying human cumulative culture. *Science*, *335*(6072), 1114–1118. https://doi.org/10.1126/science.1213969

Dean, L. G., Vale, G. L., Laland, K. N., Flynn, E., & Kendal, R. L. (2014). Human cumulative culture: A comparative perspective. *Biological Reviews*, *89*(2), 284–301. https://doi.org/10.1111/brv.12053

Derex, M., Godelle, B., & Raymond, M. (2012). Social learners require process information to outperform individual learners. *Evolution*, *67*(3), 688–697. https://doi.org/10.1111/j.1558-5646.2012.01804.x

Diamond, A. (2013). Executive functions. In *Annual Review of Psychology* (Vol. 64, pp. 135–168). Annual Reviews Inc. https://doi.org/10.1146/annurev-psych-113011-143750

Dindo, M., Thierry, B., & Whiten, A. (2008). Social diffusion of novel foraging methods in brown capuchin monkeys (*Cebus apella*). *Proceedings of the Royal Society B: Biological Sciences*, *275*(1631), 187–193. https://doi.org/10.1098/rspb.2007.1318

Dindo, M., Whiten, A., & de Waal, F. B. M. (2009). In-group conformity sustains different foraging traditions in capuchin monkeys (*Cebus apella*). *PLoS ONE*, *4*(11), e7858. https://doi.org/10.1371/journal.pone.0007858

Dunstone, J., Atkinson, M., Grainger, C., Renner, E., & Caldwell, C. A. (2021). Limited evidence for executive function load impairing selective copying in a win-stay lose-shift task. *PLOS ONE*, *16*(3), e0247183. https://doi.org/10.1371/journal.pone.0247183

Dunstone, J., & Caldwell, C. A. (2018). Cumulative culture and explicit metacognition: a review of theories, evidence and key predictions. *Palgrave Communications*, *4*(1), 145. https://doi.org/10.1057/s41599-018-0200-y

Enquist, M., Eriksson, K., & Ghirlanda, S. (2007). Critical social learning: A solution to Rogers's paradox of nonadaptive culture. *American Anthropologist*, *109*(4), 727–734. https://doi.org/10.1525/aa.2007.109.4.727

Enquist, M., Ghirlanda, S., Jarrick, A., & Wachtmeister, C.-A. (2008). Why does human culture increase exponentially? *Theoretical Population Biology*, *74*(1), 46–55. https://doi.org/10.1016/j.tpb.2008.04.007

Fagot, J., Gullstrand, J., Kemp, C., Defilles, C., & Mekaouche, M. (2014). Effects of freely accessible computerized test systems on the spontaneous behaviors and stress level of Guinea baboons (*Papio papio*). *American Journal of Primatology*, *76*(1), 56–64. https://doi.org/10.1002/ajp.22193

Fagot, J., Marzouki, Y., Huguet, P., Gullstrand, J., & Claidière, N. (2015). Assessment of social cognition in non-human primates using a network of computerized automated learning device (ALDM) test systems. *Journal of Visualized Experiments*, *2015*(99), 52798. https://doi.org/10.3791/52798

Falótico, T., Proffitt, T., Ottoni, E. B., Staff, R. A., & Haslam, M. (2019). Three thousand years of wild capuchin stone tool use. *Nature Ecology & Evolution*, *3*(7), 1034–1038. https://doi.org/10.1038/s41559-019-0904-4

Fawcett, T. W., Skinner, A. M. J., & Goldsmith, A. R. (2002). A test of imitative learning in starlings using a two-action method with an enhanced ghost control. *Animal Behaviour*, *64*(4), 547–556. https://doi.org/10.1006/anbe.2002.3092

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

Fisher, J., & Hinde, R. A. (1949). The opening of milk bottles by birds. *British Birds*, *42*, 347–357.

Fredman, T., & Whiten, A. (2008). Observational learning from tool using models by human-reared and mother-reared capuchin monkeys (*Cebus apella*). *Animal Cognition*, *11*(2), 295–309. https://doi.org/10.1007/s10071-007-0117-0

Galef, B. G. (1992). The question of animal culture. *Human Nature*, *3*(2), 157–178. https://doi.org/10.1007/BF02692251

Galef, B. G. (1995). Why behaviour patterns that animals learn socially are locally adaptive. *Animal Behaviour*, *49*(5), 1325–1334. https://doi.org/10.1006/anbe.1995.0164

Gergely, G., & Csibra, G. (2006). Sylvia's recipe: The role of imitation and pedagogy in the transmission of cultural knowledge. In N.J. Enfield & S.C. Levinson (Eds.), *Roots of human sociality: Culture, cognition, and human interaction* (pp. 229–255). Berg Publishers. https://doi.org/https://doi.org/10.4324/9781003135517

Giraldeau, L. A., Valone, T. J., & Templeton, J. J. (2002). Potential disadvantages of using socially acquired information. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 357, Issue 1427, pp. 1559–1566). Royal Society. https://doi.org/10.1098/rstb.2002.1065

Grampp, M., Sueur, C., van de Waal, E., & Botting, J. (2019). Social attention biases in juvenile wild vervet monkeys: implications for socialisation and social learning processes. *Primates*, 1–15. https://doi.org/10.1007/s10329-019-00721-4

Guida, A. (2020). *Commentary on François Osiurak and Emanuelle Reynaud's "The Elephant in the Room: What Matters Cognitively in Cumulative Technological Culture" (2019)*. https://www.journalofexpertise.org

Helfman, G. S., & Schultz, E. T. (1984). Social transmission of behavioural traditions in a coral reef fish. *Animal Behaviour*, *32*(2), 379–384. https://doi.org/10.1016/S0003-3472(84)80272-9

Henrich, J. (2004). Demography and cultural evolution: How adaptive cultural processes can produce maladaptive losses—The Tasmanian Case. *American Antiquity*, *69*(02), 197–214. https://doi.org/10.2307/4128416

Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science*, *317*(5843), 1360–1366. https://doi.org/10.1126/science.1146282

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

Heyes, C. (2016). Who knows? Metacognitive social learning strategies. *Trends in Cognitive Sciences*, *20*(3), 204–213. https://doi.org/10.1016/j.tics.2015.12.007

Heyes, C. (2018). *Cognitive Gadgets*. Harvard University Press. https://doi.org/10.4159/9780674985155

Hill, K., Barton, M., & Hurtado, A. M. (2009). The emergence of human uniqueness: Characters underlying behavioral modernity. *Evolutionary Anthropology: Issues, News, and Reviews*, *18*(5), 187–200. https://doi.org/10.1002/evan.20224

Hill, K., Walker, R. S., Božičević, M., Eder, J., Headland, T., Hewlett, B., Hurtado, A. M., Marlowe, F., Wiessner, P., & Wood, B. (2011). Co-residence patterns in hunter-gatherer societies show unique human social structure. *Science*, *331*(6022), 1286–1289. https://doi.org/10.1126/science.1199071

Hoehl, S., Keupp, S., Schleihauf, H., McGuigan, N., Buttelmann, D., & Whiten, A. (2019). ‘Over-imitation’: A review and appraisal of a decade of research. *Developmental Review*, *51*, 90–108. https://doi.org/10.1016/j.dr.2018.12.002

Hopper, L. M., Spiteri, A., Lambeth, S. P., Schapiro, S. J., Horner, V., & Whiten, A. (2007). Experimental studies of traditions and underlying transmission processes in chimpanzees. *Animal Behaviour*, *73*(6), 1021–1032. https://doi.org/10.1016/j.anbehav.2006.07.016

Hoppitt, W., & Laland, K. N. (2008). Social processes influencing learning in animals: A review of the evidence. *Advances in the Study of Behavior*, *38*, 105–165. https://doi.org/10.1016/S0065-3454(08)00003-X

Horner, V., Proctor, D., Bonnie, K. E., Whiten, A., & de Waal, F. B. M. (2010). Prestige affects cultural learning in chimpanzees. *PLoS ONE*, *5*(5), e10625. https://doi.org/10.1371/journal.pone.0010625

Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Animal Cognition*, *8*(3), 164–181. https://doi.org/10.1007/s10071-004-0239-6

Horner, V., Whiten, A., Flynn, E., & de Waal, F. B. M. (2006). Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(37), 13878–13883. https://doi.org/10.1073/pnas.0606015103

Hrubesch, C., Preuschoft, S., & Van Schaik, C. (2009). Skill mastery inhibits adoption of observed alternative solutions among chimpanzees (*Pan troglodytes*). *Animal Cognition*, *12*(2), 209–216. https://doi.org/10.1007/s10071-008-0183-y

Huffman, M. A., & Hirata, S. (2004). An experimental study of leaf swallowing in captive chimpanzees: insights into the origin of a self-medicative behavior and the role of social learning. *Primates*, *45*(2), 113–118. https://doi.org/10.1007/s10329-003-0065-5

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

Jenkins, P. F. (1978). Cultural transmission of song patterns and dialect development in a free-living bird population. *Animal Behaviour*, *26*(PART 1), 50–78. https://doi.org/10.1016/0003-3472(78)90007-6

Jesmer, B. R., Merkle, J. A., Goheen, J. R., Aikens, E. O., Beck, J. L., Courtemanch, A. B., Hurley, M. A., McWhirter, D. E., Miyasaki, H. M., Monteith, K. L., & Kauffman, M. J. (2018). Is ungulate migration culturally transmitted? Evidence of social learning from translocated animals. *Science*, *361*(6406), 1023–1025. https://doi.org/10.1126/science.aat0985

Jiménez, Á. V., & Mesoudi, A. (2020). Prestige does not affect the cultural transmission of novel controversial arguments in an online transmission chain experiment. *Journal of Cognition and Culture*, *20*(3–4), 238–261. https://doi.org/10.1163/15685373-12340083

Karadeniz, P. G., & Ercan, I. (2017). Examining tests for comparing survival curves with right censored data. *Stat Transit*, *18*(2), 311–328. https://www.ceeol.com/search/article-detail?id=587880

Kendal, J. R., Rendell, L., Pike, T. W., & Laland, K. N. (2009). Nine-spined sticklebacks deploy a hill-climbing social learning strategy. *Behavioral Ecology*, *20*(2), 238–244.

https://doi.org/https://doi.org/10.1093/beheco/arp016

Kendal, R. L., Boogert, N. J., Rendell, L., Laland, K. N., Webster, M., & Jones, P. L. (2018). Social learning strategies: Bridge-building between fields. In *Trends in Cognitive Sciences* (Vol. 22, Issue 7, pp. 651–665). Elsevier Ltd. https://doi.org/10.1016/j.tics.2018.04.003

Kenward, B., Weir, A. A. S., Rutz, C., & Kacelnik, A. (2005). Behavioural ecology: Tool manufacture by naive juvenile crows. *Nature*, *433*(7022), 121–121. https://doi.org/10.1038/433121a

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(31), 10681–10686. https://doi.org/10.1073/pnas.0707835105

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, *354*(6308), 110–114. https://doi.org/10.1126/science.aaf8110

Kühl, H. S., Kalan, A. K., Arandjelovic, M., Aubert, F., D'Auvergne, L., Goedmakers, A., Jones, S., Kehoe, L., Regnaut, S., Tickle, A., Ton, E., van Schijndel, J., Abwe, E. E., Angedakin, S., Agbor, A., Ayimisin, E. A., Bailey, E., Bessone, M., Bonnet, M., … Boesch, C. (2016). Chimpanzee accumulative stone throwing. *Scientific Reports*, *6*, 22219. https://doi.org/10.1038/srep22219

Laland, K. N. (2004). Social learning strategies. *Animal Learning & Behavior*, *32*(1), 4–14. https://doi.org/10.3758/BF03196002

Laland, K. N., & Galef, B. G. (2009). *The question of animal culture* (eds). Harvard University Press.

Laland, K. N., & Hoppitt, W. (2003). Do animals have culture? *Evolutionary Anthropology*, *12*(3), 150–159. https://doi.org/10.1002/evan.10111

Laland, K. N., & Seed, A. (2021). Understanding human cognitive uniqueness. *Annual Review of Psychology*, *72*(1), 689–716. https://doi.org/10.1146/annurev-psych-062220-051256

Legare, C. H., & Nielsen, M. (2015). Imitation and innovation: The dual engines of cultural learning. *Trends in Cognitive Sciences*, *19*(11), 688–699. https://doi.org/10.1016/j.tics.2015.08.005

Legare, C. H., Wen, N. J., Herrmann, P. A., & Whitehouse, H. (2015). Imitative flexibility and the development of cultural learning. *Cognition*, *142*, 351–361. https://doi.org/10.1016/j.cognition.2015.05.020

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). *Emmeans - Estimated marginal* *means, aka least-squares means.* https://cran.r-project.org/web/packages/emmeans/index.html

Leonardi, R., Buchanan-Smith, H. M., Dufour, V., MacDonald, C., & Whiten, A. (2010). Living together: behavior and welfare in single and mixed species groups of capuchin (*Cebus apella*) and squirrel monkeys (*Saimiri sciureus*). *American Journal of Primatology*, *72*(1), 33–47. https://doi.org/10.1002/ajp.20748

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

Lucas, A. J., Kings, M., Whittle, D., Davey, E., Happé, F., Caldwell, C. A., & Thornton, A. (2020). The value of teaching increases with tool complexity in cumulative cultural evolution. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1939), 20201885. https://doi.org/10.1098/rspb.2020.1885

Lynch, A., & Baker, A. J. (1993). A population memetics approach to cultural evolution in chaffinch song: meme diversity within populations. *American Naturalist*, *141*(4), 597–620. https://doi.org/10.1086/285493

Mackintosh, G. (2020). *The role of mental state understanding in distinctively human cumulative cultural evolution* (Unpublished doctoral dissertation). University of Stirling, Stirling, UK.

Marshall-Pescini, S., & Whiten, A. (2008). Chimpanzees (*Pan troglodytes*) and the question of cumulative culture: an experimental approach. *Animal Cognition*, *11*(3), 449–456. https://doi.org/10.1007/s10071-007-0135-y

Mendes, N., Hanus, D., & Call, J. (2007). Raising the level: orangutans use water as a tool. *Biology Letters*, *3*(5), 453–455. https://doi.org/https://doi.org/10.1098/rsbl.2007.0198

Mesoudi, A., & Thornton, A. (2018). What is cumulative cultural evolution? *Proc. R. Soc. B*, *285*(1880), 20180712. https://doi.org/10.1098/RSPB.2018.0712

Mesoudi, A., & Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1509), 3489–3501. https://doi.org/10.1098/rstb.2008.0129

Miu, E., Gulley, N., Laland, K. N., & Rendell, L. (2020). Flexible learning, rather than inveterate innovation or copying, drives cumulative knowledge gain. *Science Advances*, *6*(23), 286–291. https://doi.org/10.1126/sciadv.aaz0286

Monfardini, E., Gaveau, V., Boussaoud, D., Hadj-Bouziane, F., & Meunier, M. (2012). Social learning as a way to overcome choice-induced preferences? Insights from humans and rhesus macaques. *Frontiers in Neuroscience*, *6*, 127. https://doi.org/10.3389/fnins.2012.00127

Monfardini, E., Hadj-Bouziane, F., & Meunier, M. (2014). Model-observer similarity, error modeling and social learning in rhesus macaques. *PLoS ONE*, *9*(2), 1–8. https://doi.org/10.1371/journal.pone.0089825

Montrey, M., & Shultz, T. R. (2020). The evolution of high-fidelity social learning. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1928), 20200090. https://doi.org/10.1098/rspb.2020.0090

Morgan, T. J. H., Uomini, N. T., Rendell, L., Chouinard-Thuly, L., Street, S. E., Lewis, H. M., Cross, C. P., Evans, C., Kearney, R., de la Torre, I., Whiten, A., & Laland, K. N. (2015). Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature Communications*, *6*, 6029. https://doi.org/10.1038/ncomms7029

Nagell, K., Olguin, R. S., & Tomasello, M. (1993). Processes of social learning in the tool use of chimpanzees (*Pan troglodytes*) and human children (*Homo sapiens*). *Journal of Comparative Psychology*, *107*(2), 174–186. https://doi.org/10.1037/0735-7036.107.2.174

Nielsen, M., & Mark. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology*, *42*(3), 555–565. https://doi.org/10.1037/0012-1649.42.3.555

O'Sullivan, E. P., Claidière, N., & Caldwell, C. A. (2017). Action-matching biases in monkeys (*Sapajus* spp.) in a stimulus–response compatibility task: Evaluating experience-dependent malleability. *Journal of Comparative Psychology*, *131*(4), 337–347. https://doi.org/10.1037/com0000081

Osiurak, F., & Reynaud, E. (2019). The elephant in the room: What matters cognitively in cumulative technological culture. *Behavioral and Brain Sciences*, 1–57. https://doi.org/10.1017/s0140525x19003236

Ottoni, E. B., de Resende, B. D., & Izar, P. (2005). Watching the best nutcrackers: what capuchin monkeys (*Cebus apella*) know about others' tool-using skills. *Animal Cognition*, *8*(4), 215–219. https://doi.org/10.1007/s10071-004-0245-8

Panger, M. A., Perry, S., Rose, L., Gros-Louis, J., Vogel, E., Mackinnon, K. C., & Baker, M. (2002). Cross-site differences in foraging behavior of white-faced capuchins (*Cebus capucinus*).

American Journal of Physical Anthropology, *119*(1), 52–66. https://doi.org/10.1002/ajpa.10103

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & 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

Perry, S. (2011). Social traditions and social learning in capuchin monkeys (*Cebus*). *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *366*(1567). https://doi.org/https://doi.org/10.1098/rstb.2010.0317

Perry, S., Baker, M., Fedigan, L., Gros-Louis, J., Jack, K., MacKinnon, K. C., Manson, J. H., Panger, M., Pyle, K., & Rose, L. (2003). Social conventions in wild white-faced capuchin monkeys. *Current Anthropology*, *44*(2), 241–268. https://doi.org/10.1086/345825

Perry, S., & Rose, L. (1994). Begging and transfer of coati meat by white-faced capuchin monkeys, *Cebus capucinus*. *Primates*, *35*(4), 409–415. https://doi.org/10.1007/BF02381950

Price, E., Wood, L. A., & Whiten, A. (2017). Adaptive cultural transmission biases in children and nonhuman primates. *Infant Behavior and Development*, *48*, 45–53. https://doi.org/10.1016/j.infbeh.2016.11.003

R Core Team. (2020). *R: The R rroject for statistical computing*. https://www.r-project.org/

Rakoczy, H., Hamann, K., Warneken, F., & Tomasello, M. (2010). Bigger knows better: Young children selectively learn rule games from adults rather than from peers. *British Journal of Developmental Psychology*, *28*(4), 785–798. https://doi.org/10.1348/026151009X479178

Range, F., Viranyi, Z., Huber, L., Behne, T., Moll, H., & Csanyi, V. (2007). Selective imitation in domestic dogs. *Current Biology*, *17*(10), 868–872. https://doi.org/10.1016/j.cub.2007.04.026

Read, D. W. (2008). Working Memory: A cognitive limit to non-human primate recursive thinking prior to hominid evolution. *Evolutionary Psychology*, *6*(4), 147470490800600. https://doi.org/10.1177/147470490800600413

Reader, S. M., & Laland, K. N. (2003). Animal innovation: An introduction. In *Animal Innovation* (EDS, pp. 3–36). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198526223.003.0001

Reader, S. M., Morand-Ferron, J., & Flynn, E. (2016). Animal and human innovation: Novel problems and novel solutions. *Philosophical Transactions of the Royal Society B: Biological*

Sciences, *371*(1690), 20150182. https://doi.org/10.1098/rstb.2015.0182

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

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

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., Fogarty, L., Ghirlanda, S., Lillicrap, T., & Laland, K. N. (2010). Why copy others? Insights from the social learning strategies tournament. *Science*, *328*(5975), 208–213. https://doi.org/10.1126/science.1184719

Rendell, L., Fogarty, L., Hoppitt, W., Morgan, T. J. H., Webster, M. M., & Laland, K. N. (2011). Cognitive culture: theoretical and empirical insights into social learning strategies. *Trends in Cognitive Sciences*, *15*(2), 68–76. https://doi.org/10.1016/j.tics.2010.12.002

Rendell, L., & Whitehead, H. (2001). Culture in whales and dolphins. *Behavioral and Brain Sciences*, *24*(02), 309–324. https://doi.org/10.1017/S0140525X0100396X

Renner, E., Atkinson, M., & Caldwell, C. A. (2019). Squirrel monkey responses to information from social demonstration and individual exploration using touchscreen and object choice tasks. *PeerJ*, *7*, e7960. https://doi.org/10.7717/peerj.7960

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*(1), 1043. https://doi.org/10.1038/s41598-020-80221-4

Rogers, A. R. (1988). Does biology constrain culture? *American Anthropologist*, *90*(4), 819–831. https://doi.org/10.1525/aa.1988.90.4.02a00030

RStudio Team. (2014). *RStudio: Integrated development environment for R [Computer software].* RStudio Inc. https://www.r-project.org/

Saldana, C., Fagot, J., Kirby, S., Smith, K., & Claidière, N. (2019). High-fidelity copying is not necessarily the key to cumulative cultural evolution: a study in monkeys and children. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1904), 20190729. https://doi.org/10.1098/rspb.2019.0729

Santorelli, C. J., Schaffner, C. M., Campbell, C. J., Notman, H., Pavelka, M. S., Weghorst, J. A., & Aureli, F. (2011). Traditions in spider monkeys are biased towards the social domain. *PLoS ONE*, *6*(2), e16863. https://doi.org/10.1371/journal.pone.0016863

Sanz, C. M., Call, J., & Morgan, D. (2009). Design complexity in termite-fishing tools of chimpanzees (*Pan troglodytes*). *Biology Letters*, *5*(3), 293–296. https://doi.org/10.1098/rsbl.2008.0786

Sanz, C. M., & Morgan, D. B. (2009). Flexible and persistent tool-using strategies in honey-gathering by wild chimpanzees. *International Journal of Primatology*, *30*(3), 411–427. https://doi.org/10.1007/s10764-009-9350-5

Sanz, C. M., Morgan, D., & Gulick, S. (2004). New insights into chimpanzees, tools, and termites from the Congo Basin. *The American Naturalist*, *164*(5), 567–581. https://doi.org/10.1086/424803

Sanz, C. M., Schöning, C., & Morgan, D. B. (2010). Chimpanzees prey on army ants with specialized tool set. *American Journal of Primatology*, *72*(1), 17–24. https://doi.org/10.1002/ajp.20744

Sarabian, C., & MacIntosh, A. J. J. (2015). Hygienic tendencies correlate with low geohelminth infection in free-ranging macaques. *Biology Letters*, *11*(11), 20150757. https://doi.org/10.1098/rsbl.2015.0757

Sasaki, T., & Biro, D. (2017). Cumulative culture can emerge from collective intelligence in animal groups. *Nature Communications*, *8*(1), 1–6. https://doi.org/10.1038/ncomms15049

Schofield, D. P., McGrew, W. C., Takahashi, A., & Hirata, S. (2018). Cumulative culture in nonhumans: overlooked findings from Japanese monkeys? *Primates*, *59*(2), 113–122. https://doi.org/10.1007/s10329-017-0642-7

Shane, M. S., Stevens, M., Harenski, C. L., & Kiehl, K. A. (2008). Neural correlates of the processing of another's mistakes: A possible underpinning for social and observational learning. *NeuroImage*, *42*(1), 450–459. https://doi.org/10.1016/j.neuroimage.2007.12.067

Sherry, D. F., & Galef, B. G. (1984). Cultural transmission without imitation: Milk bottle opening by birds. *Animal Behaviour*, *32*(3), 937–938. https://doi.org/10.1016/S0003-3472(84)80185-2

Shettleworth, S. J. (2010). Clever animals and killjoy explanations in comparative psychology. *Trends in Cognitive Sciences*, *14*(11), 477–481. https://doi.org/10.1016/j.tics.2010.07.002

Smith, T. R., Parrish, A. E., Creamer, C., Rossettie, M., & Beran, M. J. (2020). Capuchin monkeys (sometimes) go when they know: Confidence movements in *Sapajus apella*. *Cognition*, *199*, 104237. https://doi.org/10.1016/j.cognition.2020.104237

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: on the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B*, *364*(1528), 2405–2415. https://doi.org/10.1098/rstb.2009.0052

Tennie, C., Call, J., & Tomasello, M. (2012). Untrained chimpanzees (*Pan troglodytes schweinfurthii*) Fail to Imitate Novel Actions. *PLoS ONE*, *7*(8), e41548. https://doi.org/10.1371/journal.pone.0041548

Tennie, C., Hedwig, D., Call, J., & Tomasello, M. (2008). An experimental study of nettle feeding in captive gorillas. *American Journal of Primatology*, *70*(6), 584–593. https://doi.org/10.1002/ajp.20532

Thornton, A., & McAuliffe, K. (2006). Teaching in wild meerkats. *Science*, *313*(5784), 277–229. https://doi.org/10.1126/science.1128727

Tomasello, M. (1990). Cultural transmission in the tool use and communicatory signaling of chimpanzees? In S. T. Gibson & K. R. Parker (Eds.), *“Language” and Intelligence in Monkeys and Apes: Comparative Developmental Perspectives* (pp. 274–311). Cambridge University Press. https://doi.org/doi:10.1017/CBO9780511665486.012

Tomasello, M. (1999a). *The cultural origins of human cognition* (vi). Harvard University Press.

Tomasello, M. (1999b). The human adaptation for culture. *Annual Review of Anthropology*, *28*(1), 509–529. https://doi.org/10.1146/annurev.anthro.28.1.509

Tomasello, M., Davis-Dasilva, M., Camak, L., & Bard, K. (1987). Observational learning of tool-use by young chimpanzees. *Human Evolution*, *2*(2), 175–183. https://doi.org/10.1007/BF02436405

Vale, G. L., McGuigan, N., Burdett, E., Lambeth, S. P., Lucas, A., Rawlings, B., Schapiro, S. J., Watson, S. K., & Whiten, A. (2020). Why do chimpanzees have diverse behavioral repertoires yet lack more complex cultures? Invention and social information use in a cumulative task. *Evolution and Human Behavior*. https://doi.org/10.1016/j.evolhumbehav.2020.11.003

van der Post, D. J., Franz, M., & Laland, K. N. (2017). The evolution of social learning mechanisms and cultural phenomena in group foragers. *BMC Evolutionary Biology*, *17*(1), 1–15. https://doi.org/10.1186/s12862-017-0889-z

Van Leeuwen, E. J. C., Cronin, K. A., Schütte, S., Call, J., & Haun, D. B. M. (2013). Chimpanzees (*Pan* *troglodytes*) flexibly adjust their behaviour in order to maximize payoffs, Not to Conform to Majorities. *PLoS ONE*, *8*(11), e80945. https://doi.org/10.1371/journal.pone.0080945

van Schaik, C. P., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C. D., Singleton, I., Suzuki, A., Utami, S. S., & Merrill, M. (2003). Orangutan cultures and the evolution of material culture. *Science*, *299*(5603), 102–105. https://doi.org/10.1126/science.1078004

Visalberghi, E., & Fragaszy, D. M. (1990). “Language” and intelligence in monkeys and apes. In S. T. Parker & K. R. Gibson (Eds.), *"Language” and intelligence in monkeys and apes: Comparative developmental perspectives* (pp. 247–273). Cambridge University Press. https://doi.org/10.1017/CBO9780511665486

Voelkl, B., & Huber, L. (2000). True imitation in marmosets. *Animal Behaviour*, *60*(2), 195–202. https://doi.org/10.1006/anbe.2000.1457

Waitt, C., & Buchanan-Smith, H. M. (2006). Perceptual considerations in the use of colored photographic and video stimuli to study nonhuman primate behavior. *American Journal of Primatology*, *68*(11), 1054–1067. https://doi.org/10.1002/ajp.20303

Wasielewski, H. (2014). Imitation Is necessary for cumulative cultural evolution in an unfamiliar, opaque task. *Human Nature*, *25*(1), 161–179. https://doi.org/10.1007/s12110-014-9192-5

Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C. E. G., Wrangham, R. W., & Boesch, C. (1999). Cultures in chimpanzees. *Nature*, *399*(6737), 682–685. https://doi.org/10.1038/21415

Whiten, A., & Ham, R. (1992). On the nature and evolution of imitation in the animal kingdom: Reappraisal of a century of research. *Advance in the Study of Behaviour*, *21*, 239–283. https://doi.org/10.1016/S0065-3454(08)60146-1

Whiten, A., McGuigan, N., Marshall-Pescini, S., & Hopper, L. M. (2009). Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2417–2428. https://doi.org/10.1098/rstb.2009.0069

Whiten, A., & van Schaik, C. P. (2007). The evolution of animal ‘cultures’ and social intelligence. *Philosophical Transactions of the Royal Society B*, *362*, 603–620. https://doi.org/10.1111/J.1467-9280.1996.TB00386.X

Wilks, C. E. H., Atkinson, M., & Caldwell, C. A. (under revision). Children’s use of social information from multiple models: cognitive capacities underlying population size effects on cumulative culture. *Culture and Evolution*.

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*, *204*, 105031. https://doi.org/10.1016/j.jecp.2020.105031

Wood, L. A., & Whiten, A. (2017). Visible spatial contiguity of social information and reward affects social learning in brown capuchins (*Sapajus apella*) and children (*Homo sapiens*). *Journal of Comparative Psychology*, *131*(4), 304–316. https://doi.org/10.1037/com0000076

Wynn, T., & Coolidge, F. L. (2007). Did a small but significant enhancement in working memory capacity power the evolution of modern thinking. In P. Mellars, K. Boyle, O. Bar-Yosef, & C. Stringer (Eds.), *Rethinking the Human Revolution.* (pp. 79–90). McDonald Institute Monographs.

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

Yamamoto, S., Humle, T., & Tanaka, M. (2013). Basis for cumulative cultural evolution in chimpanzees: social learning of a more efficient tool-use technique. *PLoS ONE*, *8*(1), e55768. https://doi.org/10.1371/journal.pone.0055768

Yu, R., & Zhou, X. (2006). Brain responses to outcomes of one's own and other's performance in a gambling task. *NeuroReport*, *17*(16), 1747–1751. https://doi.org/10.1097/01.wnr.0000239960.98813.50

Zwirner, E., & Thornton, A. (2015). Cognitive requirements of cumulative culture: teaching is useful but not essential. *Scientific Reports*, *5*, 16781. https://doi.org/10.1038/srep16781

Appendices

Appendix 1. Children PFR Task Script for Chapter 3.

Introduction

- *I have a game about the Wild West where your job is to find as many wagon wheels as you can! And you're going to play the game with Handy Andy. Andy will pick some boxes first and some of them will have wagon wheels in and some won't. Then it will be your turn and your job is to find as many wagon wheels as possible.*

Block 1 Practise

- *Here's a practise turn. Andy will have a go first.*
- *Look, Andy found 1 wagon wheel and 1 empty box.*
- *Now it's your turn to try to find all the wagon wheels. Can you press the green continue button?*
    - [if transient first]: *you have to remember what Andy found.*
      [If visible first]: *You can see what Andy found.*
- *Now you can pick some boxes!*
- *You found (1 / 2) wagon wheels, so you get (1 / 2) marbles*. [put that number of marbles in their box]
- *If you get this many marbles, you get one of these stickers* [show small sticker]. *If you get this many marbles you get one of these stickers* [show large sticker].
- *That was a practise. Would you like to keep playing?*

Block 1 Test Trials

- *Handy Andy has a go first.*
    - *He found X wagon wheels.*
- *Okay, now it's your turn.*
    - *Great job, you found X wagon wheels! You get X marbles.*
      [or] *No wagon wheels! Let's see if you can find more next time.*
- *Let's try another one.*

- [Repeat]

Block 2 Intro

- *It's almost the same now but this time:*

  [visible]: *you can see what Andy found*.

  [transient]: *you have to try to remember what Handy Andy found.*

Block 2 Practise

- *Let's have a practise go.*
  - *Handy Andy found 1 wagon wheel and 1 empty box!*
- *Now it's your go, can you press the green button?*
- *This time:*

  [visible] *you can still see how Andy did.*

  [transient] *you can't see what Andy found this time so you will need to remember.*
  - *Okay, can you find the wagon wheels?*
  - *Well done you found (1 / 2) wagon wheels.*
- *Are you ready to keep playing?*

Block 2 Test Trials

- [same as Block 1 test trials]

End

- *"Well done you got lots of marbles! Good Job"* [give sticker]

Appendix 2. Average task scores (and standard deviations (SD)) overall and grouped by demonstration score (DS), and correlations and t-test results for Chapter 3. *Separated by age bracket and cue type. Green highlighting indicates values that achieved the associated PFR Classification criteria (Table 3.2).*

3a. Stage C: 9-stimuli

Age	**Cue type**	**Overall Score (chance = 1/3)**				**Correlation**		**DS = 1**		**DS = 2**		**Highest PFR level met**
		Mean (/3)	*SD*	*t*	*p*	*r*	*p*	*Mean (/3)*	*SD*	*Mean (/3)*	*SD*	
2-3	transient	1.35	0.84	3.14	<.001*	0.06	.48	1.27	0.86	1.53	0.73	**1**
4	transient	1.70	0.90	6.54	<.001*	0.46	<.001*	1.44	0.50	1.81	0.89	**2**
5	transient	2.00	0.90	7.71	<.001*	0.74	<.001*	1.40	0.81	2.30	0.47	**3**
6-7	transient	1.88	0.83	4.78	<.001*	0.58	<.001*	1.62	0.49	2.08	0.27	**3**
2-3	visible	1.58	0.97	5.33	<.001*	0.55	<.001*	1.25	0.44	1.81	0.39	**2**
4	visible	1.69	1.01	6.10	<.001*	0.52	<.001*	1.31	0.47	2.07	0.60	**3**
5	visible	1.90	0.88	6.68	<.001*	0.49	<.001*	1.40	0.81	1.90	0.71	**2**
6-7	visible	1.94	0.87	5.10	<.001*	0.52	<.001*	1.46	0.51	2.00	0.40	**2**

3b. Stage D: 16-stimuli

Age	**Cue type**	**Overall Score (chance = 1/4)**				**Correlation**		**DS = 1**		**DS = 2**		**DS = 3**		**Highest PFR level met**
		Mean (/4)	*SD*	*t*	*p*	*r*	*p*	*Mean (/4)*	*SD*	*Mean (/4)*	*SD*	*Mean (/4)*	*SD*	
2-3	transient	1.31	0.91	2.83	.002*	0.06	.32	1.33	0.71	1.20	0.76	1.53	1.10	**1**
4	transient	1.73	1.12	6.54	<.001*	0.48	<.001*	1.38	0.79	1.56	0.87	2.12	1.00	**2**
5	transient	2.10	1.03	7.77	<.001*	0.35	<.001*	1.60	0.67	2.10	0.55	2.30	1.02	**3**
6-7	transient	2.23	1.15	6.43	<.001*	0.66	<.001*	1.69	0.61	2.31	0.47	2.77	0.43	**3**
2-3	visible	1.95	1.31	8.49	<.001*	0.63	<.001*	1.25	0.67	2.06	0.43	2.31	0.99	**3**
4	visible	2.03	1.32	8.88	<.001*	0.64	<.001*	1.40	0.72	1.88	0.60	2.67	0.88	**2**
5	visible	2.41	1.07	10.04	<.001*	0.45	<.001*	1.89	0.58	2.10	0.96	2.80	0.61	**3**
6-7	visible	2.26	1.18	7.34	<.001*	0.55	<.001*	1.69	0.73	2.15	0.54	2.85	0.87	**3**

Appendix 3. Number of PFR training trials on visible cues for Chapter 4. *Green highlighting indicates criterion was met.*

a. Edinburgh

Subject	*Alba*	*Bear*	*Carlos*	*Chico*	*Flojo*	*Fudge*	*Hazel*	*Inti*	*Junon*	*Kato*	*Luna*	*Mekoe*	*Pedra*	*Pixie*	*Reuben*	*Rosa*	*Rufo*	*Torres*	*Ximo*
Stage A	335	372	1577	931	173	32	408	556	1188	1104	995	416	62	333	960	189	463	427	505
Stage B	63	318	405	630	383	*NA*	910	1162	315	63	357	*NA*	*NA*	*NA*	756	*NA*	237	1214	1056
Stage C	113	*NA*	*NA*	*NA*	60	*NA*	*NA*	*NA*	*NA*	60	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	120	*NA*	60
Stage D	218	*NA*	*NA*	*NA*	60	*NA*	*NA*	*NA*	*NA*	360	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	60	*NA*	*NA*

b. Kyoto

Subject	*Heiji*	*Kojilo*	*Pigmon*	*Zen*	*Zephi*	*Zinnia*
Stage A	280	700	940	338	460	1600
Stage B	419	540	60	360	660	884
Stage C	60	60	60	120	60	120
Stage D	120	60	60	60	60	60

Appendix 4. Average task scores (and standard deviations (SD)) overall and grouped by demonstration score (DS), and correlations and t-tests results. Chapter 4. *Overall population, and individual subject results are presented. Green highlighting indicates values that achieved the associated PFR Classification criteria (Table 3.2).*

a. *Stage C: 9-stimuli*

Population/Subject		**Overall Score (chance = 1/3)**				**Correlation**		**DS = 1**				**DS = 2**				**Highest PFR level**
		Mean (/3)	*SD*	*t*	*p*	*r*	*p*	*Mean (/3)*	*SD*	*t*	*p*	*Mean (/3)*	*SD*	*t*	*p*	*(numerical/ significant)*
	Edinburgh	1.63	0.73	13.82	<.001*	0.26	<.001*	1.54	0.66	5.96	<.001*	1.71	0.70	NA	NA	**2**
	Kyoto	1.74	0.78	24.93	<.001*	0.50	<.001*	1.42	0.63	6.99	<.001*	1.97	0.57	NA	NA	**2**
Alba	Edi	1.44	0.76	4.51	<.001*	0.17	.20	1.45	0.51	2.31	.01*	1.45	1.00	NA	NA	**1**
Flojo	Edi	1.70	0.65	7.64	<.001*	0.17	.20	1.64	0.79	3.44	<.001*	1.79	0.51	NA	NA	**1**
Kato	Edi	1.59	0.69	6.07	<.001*	0.45	<.001*	1.40	0.60	2.05	.02*	1.80	0.41	NA	NA	**2**
Rufo	Edi	1.68	0.77	7.70	<.001*	0.31	.009*	1.61	0.66	3.36	<.001*	1.68	0.75	NA	NA	**2**
Ximo	Edi	1.95	0.83	5.10	<.001*	0.24	.40	1.80	0.84	2.04	.03*	2.20	0.45	0.56	.29	**3 / 1**
Heiji	Kyo	1.63	0.76	8.03	<.001*	0.32	.002*	1.33	0.61	2.11	.02*	1.73	0.74	NA	NA	**2**
Kojilo	Kyo	1.97	0.79	12.85	<.001*	0.66	<.001*	1.53	0.63	3.36	<.001*	2.13	0.35	0.93	.18	**3 / 2**
Pigmon	Kyo	1.82	0.82	10.55	<.001*	0.50	<.001*	1.53	0.73	3.36	<.001*	2.17	0.53	1.17	.12	**3 / 2**
Zen	Kyo	1.60	0.78	10.32	<.001*	0.53	<.001*	1.33	0.55	2.92	.002*	1.89	0.49	NA	NA	**2**
Zephi	Kyo	1.88	0.71	11.54	<.001*	0.62	<.001*	1.50	0.57	3.15	<.001*	2.07	0.45	.45	.33	**3/2**
Zinnia	Kyo	1.68	0.78	8.70	<.001*	0.46	<.001*	1.37	0.72	2.32	.01*	1.90	0.71	NA	NA	**2**

b. Stage D: 16-stimuli

Population/Subject		**Overall Score (chance = 1/4)**				**Correlation**		**DS = 1**				**DS = 2**			
		Mean (/4)	*SD*	*t*	*p*	*r*	*p*	*Mean (/4)*	*SD*	*t*	*p*	*Mean (/4)*	*SD*	*t*	*p*
	Edinburgh	1.95	0.91	19.29	<.001*	0.47	<.001*	1.57	0.69	5.23	<.001*	2.08	0.72	0.77	.22
	Kyoto	2.18	1.02	55.90	<.001*	0.64	<.001*	1.61	0.70	13.15	<.001*	2.11	0.67	2.38	.008*
Flojo	Edi	1.88	0.84	8.01	<.001*	0.49	<.001*	1.45	0.51	2.08	.02*	1.80	0.89	NA	NA
Kato	Edi	2.06	0.90	15.41	<.001*	0.50	<.001*	1.76	0.70	4.86	<.001*	2.21	0.60	1.38	.09*
Rufo	Edi	1.79	0.89	7.93	<.001*	0.43	<.001*	1.30	0.73	1.42	.08*	2.10	0.72	0.45	.33
Heiji	Kyo	2.06	0.90	25.93	<.001*	0.46	<.001*	1.69	0.67	7.66	<.001*	2.08	0.66	0.82	.21
Kojilo	Kyo	2.18	1.12	18.40	<.001*	0.69	<.001*	1.52	0.80	3.71	<.001*	2.15	0.82	1.01	.16
Pigmon	Kyo	2.33	1.07	27.70	<.001*	0.80	<.001*	1.62	0.71	5.77	<.001*	2.13	0.51	1.20	.12
Zen	Kyo	2.21	0.95	26.71	<.001*	0.64	<.001*	1.70	0.62	6.91	<.001*	2.28	0.61	2.78	.003*
Zephi	Kyo	2.17	1.13	22.26	<.001*	0.69	<.001*	1.43	0.77	3.81	<.001*	1.96	0.83	NA	NA
Zinnia	Kyo	2.17	1.06	12.91	<.001*	0.75	<.001*	1.58	0.58	2.91	.002*	1.96	0.46	NA	NA

Population/Subject		**DS = 3**				**Highest PFR level**
		Mean (/4)	*SD*	*t*	*p*	*(numerical/significant)*
	Edinburgh	2.25	0.71	NA	NA	**3 / 2**
	Kyoto	2.72	0.68	NA	NA	**3**
Flojo	Edi	2.10	0.79	NA	NA	**2**
Kato	Edi	2.47	0.63	NA	NA	**3**
Rufo	Edi	1.95	0.69	NA	NA	**3 / 2**
Heiji	Kyo	2.48	0.73	NA	NA	**3 / 2**
Kojilo	Kyo	2.75	0.76	NA	NA	**3 / 2**
Pigmon	Kyo	2.96	0.42	NA	NA	**3 / 2**
Zen	Kyo	2.80	0.54	NA	NA	**3**
Zephi	Kyo	2.68	0.85	NA	NA	**2**
Zinnia	Kyo	2.83	0.48	NA	NA	**2**

Appendix 5. Number of PFR training trials on transient cues for Chapter 4. *Green highlighting indicates criterion was met.*

Subject	*Heiji*	*Kojilo*	*Pigmon*	*Zen*	*Zephi*	*Zinnia*
Stage A	2100	2280	4260	3660	2100	1980
Stage B	1988	1920	*NA*	*NA*	1802	960
Stage C	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
Stage D	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*

Appendix 6. Number of PFR training problems for Chapter 5. *Green highlighting indicates criterion was met. Grey highlighting indicates subjects/stages affected by the technical issue described in the Results section.*

Subject	Angele	Arielle	Articho	Atmosp-here	Bobo	Brigitte	Cauet	Dora	Dream	Ewine	Fana	Felipe	Feya	Flute	Harlem	Hermine	Kali
Stage A	120	120	120	360	360	180	240	120	120	120	120	120	120	120	300	120	180
Stage B	60	2160	1500	2520	*NA*	1860	1080	4320	240	600	120	660	60	300	180	1860	1080
Stage C	60	60	60	60	*NA*	60	60	60	60	60	60	60	60	60	120	60	60
Stage D	60	1681	2372	1020	*NA*	2703	240	660	300	181	720	180	661	60	60	60	3230

Subject	Lips	Lome	Mako	Mali	Muse	Petoulette	Pipo	Violette
Stage A	180	180	1320	180	480	480	480	120
Stage B	1140	2820	7980	960	5760	3027	401	120
Stage C	240	120	120	180	180	*NA*	*NA*	60
Stage D	9522	1860	2651	2580	3901	*NA*	*NA*	480

Appendix 7. Average task scores (and standard deviations (SD)) overall and grouped by demonstration score (DS), and correlation and t-test results for Chapter 5. *Overall group, and individual subject results are presented. Green highlighting indicates values that achieved the associated PFR Classification criteria (Table 3.2).*

a. Stage C: 9-stimuli

Subject	**Overall Score (chance = 1/3)**				**Correlation**		**DS = 1**				**DS = 2**				**Highest PFR level**
	Mean (/3)	*SD*	*t*	*p*	*r*	*p*	*Mean (/3)*	*SD*	*t*	*p*	*Mean (/3)*	*SD*	*t*	*p*	*(numerical/significant)*
Overall	1.62	0.87	27.32	<.001*	0.53	<.001*	1.30	0.61	6.57	<.001*	1.84	0.64	NA	NA	**2**
Angele	1.55	0.81	5.67	<.001*	0.41	.001*	1.20	0.52	1.05	.15	1.75	0.44	NA	NA	**2**
Arielle	1.76	0.88	7.98	<.001*	0.76	<.001*	1.35	0.49	1.80	.04*	2.00	0.56	NA	NA	**2**
Articho	1.54	0.95	5.54	<.001*	0.54	<.001*	1.10	0.45	0.53	.30	2.00	0.73	NA	NA	**2**
Brigitte	1.68	0.73	7.01	<.001*	0.49	<.001*	1.40	0.75	2.05	.02*	2.00	0.32	NA	NA	**2**
Cauet	1.69	0.84	7.14	<.001*	0.72	<.001*	1.25	0.44	1.30	.10	1.95	0.51	NA	NA	**2**
Dora	1.60	0.88	6.20	<.001*	0.43	<.001*	1.40	0.68	2.05	.02*	1.75	0.79	NA	NA	**2**
Dream	1.61	1.05	6.33	<.001*	0.79	<.001*	1.25	0.55	1.30	.10	1.95	0.60	NA	NA	**2**
Ewine	1.61	0.93	6.73	<.001*	0.73	<.001*	1.30	0.66	1.55	.06	1.95	0.39	NA	NA	**2**
Fana	1.39	0.72	4.01	<.001*	0.36	.004*	1.15	0.67	0.79	.22	1.20	0.77	NA	NA	**2**
Felipe	1.76	0.90	7.98	<.001*	0.78	<.001*	1.15	0.49	0.79	.22	2.00	0.65	NA	NA	**2**
Feya	1.76	0.83	6.46	<.001*	0.36	.005*	1.60	0.68	3.08	.002*	1.95	0.51	NA	NA	**2**
Flute	1.76	0.87	7.56	<.001*	0.65	<.001*	1.35	0.49	1.80	.04*	2.00	0.56	NA	NA	**2**
Harlem	1.76	0.89	6.33	<.001*	0.37	.004*	1.30	0.66	1.55	.06	1.65	0.81	NA	NA	**2**
Hermine	1.71	0.89	7.42	<.001*	0.71	<.001*	1.25	0.64	1.30	.10	1.85	0.49	NA	NA	**2**
Kali	1.35	0.83	3.63	<.001*	0.33	.01*	1.15	0.75	0.79	.22	1.55	0.69	NA	NA	**2**
Mako	1.68	0.82	7.01	<.001*	0.54	<.001*	1.40	0.68	2.05	.02*	1.65	0.67	NA	NA	**2**
Muse	1.48	0.80	4.90	<.001*	0.33	.01*	1.35	0.81	1.80	.04*	1.65	0.67	NA	NA	**2**
Violette	1.79	0.84	8.27	<.001*	0.58	<.001*	1.40	0.50	2.05	.02*	2.20	0.52	1.16	0.13	**3/2**

b. Stage D: 16-stimuli

Subject	**Overall Score (chance = 1/4)**				**Correlation**		**DS = 1**			
	Mean (/4)	*SD*	*t*	*p*	*r*	*p*	*Mean (/4)*	*SD*	*t*	*p*
Overall	1.94	1.06	32.52	<.001*	0.59	<.001*	1.38	0.72	6.11	<.001*
Angele	2.07	1.01	10.69	<.001*	0.72	<.001*	1.45	0.69	2.08	.02*
Arielle	2.02	1.10	10.19	<.001*	0.75	<.001*	1.20	0.62	0.97	0.17
Cauet	1.75	1.04	7.55	<.001*	0.65	<.001*	1.20	0.70	0.97	0.17
Dora	1.71	1.04	7.17	<.001*	0.47	<.001*	1.25	0.85	1.20	.12
Dream	1.69	1.05	6.98	<.001*	0.58	<.001*	1.05	0.60	0.25	.40
Ewine	1.89	1.02	8.90	<.001*	0.64	<.001*	1.30	0.57	1.42	.08
Felipe	2.13	1.13	11.31	<.001*	0.65	<.001*	1.35	0.59	1.64	.052
Feya	1.96	1.01	9.59	<.001*	0.50	<.001*	1.75	0.55	3.36	<.001*
Flute	2.00	1.03	9.99	<.001*	0.70	<.001*	1.45	0.69	2.08	.02*
Harlem	1.72	1.06	7.26	<.001*	0.39	<.001*	1.20	1.06	0.97	.17
Hermine	2.18	1.12	11.83	<.001*	0.64	<.001*	1.45	0.76	2.08	.02*
Violette	2.18	0.92	11.41	<.001*	0.49	<.001*	1.85	0.49	3.79	<.001*

b. Stage D continued

Subject	**DS = 2**				**DS = 3**				**Highest PFR level**
	Mean (/4)	*SD*	*t*	*p*	*Mean (/4)*	*SD*	*t*	*p*	*(numerical/significant)*
Overall	2.02	0.66	0.32	.37	2.55	0.75	NA	NA	**3/2**
Angele	2.10	0.55	0.45	.33	2.60	0.75	NA	NA	**3/2**
Arielle	2.15	0.59	0.67	.25	2.70	0.57	NA	NA	**3/2**
Cauet	1.80	0.63	NA	NA	2.20	0.77	NA	NA	**2**
Dora	1.95	0.69	NA	NA	2.20	1.01	NA	NA	**2**
Dream	1.85	0.81	NA	NA	2.20	0.95	NA	NA	**2**
Ewine	2.05	0.60	.22	.41	2.60	0.68	NA	NA	**3/2**
Felipe	2.05	0.94	.22	.41	2.85	0.75	NA	NA	**3/2**
Feya	2.10	0.55	.45	.33	2.65	0.59	NA	NA	**3/2**
Flute	2.00	0.56	NA	NA	2.85	0.37	NA	NA	**2**
Harlem	1.95	0.69	NA	NA	2.35	0.75	NA	NA	**2**
Hermine	2.20	0.62	0.89	.19	2.95	0.60	NA	NA	**3/2**
Violette	2.05	0.69	.22	.41	2.50	0.61	NA	NA	**3/2**