THE ROLE OF MENTAL STATE UNDERSTANDING IN DISTINCTIVELY HUMAN CUMULATIVE CULTURAL EVOLUTION

Gemma Mackintosh

Thesis submitted for the degree of Doctor of Philosophy

Psychology Division

University of Stirling

December 2020

For Alla, buey

Abstract

This thesis aimed to contribute to the existing literature exploring distinctively human cognitive mechanisms. Specifically, the aim was to investigate whether, and how, the distinctively human propensity for understanding the mental states of others (referred to commonly in the literature as '*theory of mind*') facilitates cumulative cultural evolution. The general methodology used throughout this thesis involved grid search tasks in which participants searched for stimuli using vicarious information generated from a participant who had already attempted the same grid search. The first experimental chapter in this thesis explored the suitability of the grid search task for capturing search behaviour in response to vicarious information about search outcomes. The second experimental chapter explored adult transmission behaviour, and whether small amounts of intentionally produced information could facilitate cumulative culture relative to small amounts of inadvertently produced information. This methodology was extended to a sample of children in Chapter 4 in order to assess whether the ability to intentionally select beneficial information to facilitate cumulative culture increases with age. The final experimental chapter explored a similar task context, but instead of manipulating downward transmissions, manipulated upward transmissions to assess whether feedback from successors influences the quality of information sent by the predecessor. Together these studies explored the ability (which may be distinctive to humans) to tailor transmitted information to the needs of a specific receiver in order to best facilitate the retention of beneficial knowledge.

This thesis found that the sharing of intentionally selected knowledge is sufficient for generating cumulative cultural evolution over generations, relative to circumstances where only inadvertent cues about a predecessor's performance is available. Furthermore, the developmental trajectory found in this capacity suggests that it may be supported by distinctively human cognitive mechanisms. We believe that capacities for understanding others' minds were responsible for the successful performance of the adults and older children in the transmission chain tasks, and we argue for the logical plausibility of this interpretation. However, other alternative interpretations of our results remain possible, and these are also discussed, along with potential future research ideas which might differentiate between competing explanations.

Acknowledgements

Thank you to Glasgow Science centre and the students of the University of Stirling and to Eoin O' Sullivan, without whom the data collection for my adult-participant chapters would have been impossible. Data collection for Chapter 5 was made possible by The Highland Council. Thanks ever so much to all the children who took part and their teachers who let me borrow them. Special thanks to Eva Catherine for keeping me right at every turn.

I am so very thankful for my fellow lab group members, Mark Atkinson and Liz Renner, for continued support, and for gifting me with skills I once thought were beyond my capability. A million thanks to Kirsten Blakey, Donna Kean, Sarah Kraemer, Juliet Robertstone and Charlotte Wilks, my PhD buddies. It's been an honour to take on this journey with you all. Thank you to the European Research Council for funding us.

I am especially grateful to my supervisor, Prof Christine Caldwell, for pointing me in the right directions; and for years of guidance and positivity. Thank you for inspiring me to be a better researcher.

My love for science exists because of my Uncle Eoin, who taught me to ask interesting questions about the world. Thank you for leading me on this path from a young age.

I would like to thank my friends, Rebecca Skinner, Jordan Miller, Lizzie Collins and Sophia Daoudi for encouragement and advice. Thank you also to my best Judy, Susan Hackett, for helping me with data collection, and to 'wee Erin' for being my consultant 10-year-old.

Eternal thanks to my partner, Tommy Conboy, for without him I would never have had the confidence to embark on this adventure. And finally, I thank my family. Words cannot express the gratitude I have for my parents, Linda and Fraser, my lovely brother Andrew, my sisters Runa and Fenn, and my grandparents, for their continued support and for always being on the other end of the phone to share my excitement.

Table of Contents

Clarification of terminologies used

Throughout this thesis I will use several terms to describe the roles of the individuals taking part in these tasks and the direction of communicative transmissions between them.

In all cases, any pair of participants who interact directly with one another include an experienced individual (who has completed their own attempt at the task), and a naïve individual (who has not yet completed their own attempt at the task). These members represent the predecessor and successor (respectively) in a cultural chain. To make these terms clear and consistent for all chapters, I will refer to the experienced predecessor as the '*cultural parent'*, and I will refer to the naïve successor as the '*cultural offspring'*. In circumstances where information from a cultural parent is not generated by a human participant (e.g., in Chapter 2 where information was computer-generated rather than using real participant data), I will refer to this as the 'parent model'.

The direction of transmission (particularly relevant for Chapter 5) refers to whether a transmission is made from the cultural parent to the cultural offspring, or vice versa. Information transmitted from cultural parent to cultural offspring is referred to as '*downward transmission'*, and information transmitted from cultural offspring to cultural parent is referred to as '*upward transmission'*.

The terms '*intentional knowledge sharing'*, '*intentional knowledge transmission'*, *intentional information sharing', 'intentional information transmission'* and '*teaching'* are all used, relatively interchangeably, to mean the deliberate sending of known information to a cultural offspring. Use of these terms was decided based on the appropriateness to the context each time.

Finally, throughout this thesis I refer to '*mental state understanding'*, however, similar terms ('*theory of mind', 'mental state inference', 'mentalizing', 'mindreading'* etc.) are also used commonly throughout the literature to mean the same thing. In some context-dependant cases, however, some of the other terms are used within this thesis.

Chapter 1: General Introduction

1.1 Opening Remarks

The purpose of this review is to explain how human capacities for understanding the mental states of others could potentially help us to understand key differences between human and non-human culture. This chapter will argue for a facilitatory effect of mental state understanding in distinctively human cumulative culture, and briefly introduce the methods used to test this theory in the following chapters of this thesis. First, I will discuss the evidence for cumulative culture, and its supporting capacities, in non-humans, as well as pinpointing what about it is specifically unique to humans. Then I will discuss the uniqueness of mental state understanding to humans, including the emergence of mental state understanding (as measured using '*elicited'* responses) in human development, and claims of capacities present in non-humans and infants (identified using '*spontaneous'* response methods). This section will go on to discuss the arguments that refute the claims of mental state understanding based on the spontaneous measures, and will argue that it is specifically the mechanisms of metarepresentational understanding that emerge around age 4 (consistent with results using elicited responses) that support cumulative culture. This theme will be continued in the following section with a discussion of the mechanisms directly supporting cumulative culture (flexible use of social learning strategies; flexible teaching; intentional bidirectional communication) and how the use of these mechanisms is made possible by maintaining and comparing representations of other minds. Finally, I will discuss the current experimental methods used to examine cumulative culture in a lab setting, and how use of an abstract, computer-based task will be used to investigate the link between mental state understanding and cumulative culture in the following chapters of this thesis.

1.2 The extent to which cumulative cultural evolution is unique to humans

Humans are the only species that unambiguously exhibit '*cumulative cultural evolution*' - a phenomenon that produces the accumulation of beneficial traits including behaviours and artefacts, such that they become more effective or efficient over time (Caldwell & Millen, 2008b; Dean et al., 2014). This is achieved when traits are cycled through repeated social transmissions over generations of learners, with adaptive, beneficial traits being retained, and maladaptive traits being pruned. It is very common for humans to generate technology and complex systems that could only be the result of repeated transmission (see Basalla (1988)). Even seemingly early artefacts of human culture have been found to have become more complex and multipurpose as a result of cumulative culture, as evidenced from the evolution of Oldowan stone tools (Lycett & Von Cramon-Taubadel, 2013).

The retention and accumulation of adaptive traits (the '*ratchet effect'*, (Tennie, Call, & Tomasello, 2009; Tomasello, 1999)) is a key characteristic of cumulative culture that allows improvement of performance over generations. As such, many generations of transmission ultimately lead to traits that are far beyond the creative abilities of a single generation (Tomasello et al., 1993). In some cases, such accumulation can lead to the origins and underlying mechanisms becoming opaque to later generations (Muthukrishna & Henrich, 2016).

Many non-human species have been found to socially transmit information within populations, leading to distinct community variations in cultural traits. Examples include geographically distinct dialects in yellowhammers (Podos & Warren, 2007), and whale song (Garland et al., 2011), as well as variations in the use of tools between crow populations (Holzhaider et al., 2010), and foraging techniques and grooming in chimpanzee populations (Whiten et al., 1999; Whiten & Van Schaik, 2007). However, despite the commonality of culture across species, the evidence of *cumulative* culture in any non-human species is questionable.

There have, however, been some claims of cumulative culture in non-humans, derived from evidence of human-like transmission and trait emergence in a variety of non-human primate species. Examples include experimental studies of chimpanzees (Vale et al., 2017) and baboons (Claidière et al., 2014), and naturalistic observations of Japanese macaques (Schofield et al., 2018). There is also supporting evidence from naturalistic observations in the non-primate literature, for example, Hunt & Gray (2003) found that New-Caledonian crows were capable of developing complex foraging tools by adapting earlier designs to increase durability. In this study, the social transmission of the tool designs was not empirically tested, and rather only implied by the design's uptake by the wider population of birds. However, empirical evidence has been found in other species. For example, social transmission leading to increased knowledge of beneficial foraging areas has been found to influence migratory behaviour of wild ungulates such as bighorn sheep and moose (Jesmer et al., 2018).

The above examples all contribute to the growing literature that both humans and non-humans satisfy the basic criterion for cumulative culture, i.e., the transmission and accumulation of beneficial traits over generations. However, there is still a general consensus that cumulative culture takes a distinctive form in humans, leading to an increased complexity of human cultures in comparison (Boyd & Richerson, 1996). Reasons proposed for this advanced cumulative culture in humans have generally focused on social learning mechanisms that may also be unique, thus opening up access to sources of information that are only available if these unique mechanisms are also in place. However, there is growing speculation as to whether these proposed supporting capacities really are unique to humans, which calls into question their proposed role as sufficient prerequisites for cumulative culture.

Imitation

Imitation (which here, consistent with other literature in this field, is defined as action-copying) occurs when an individual is able to perform a new behaviour as a result of seeing another individual perform that behaviour (Heyes, 1993). Copying the actions of others serves to facilitate the acquisition of new skills, and doing so selectively based on the outcome of other's actions (i.e., copying only successful actions of others) has been argued to form the basis for cumulative culture (Tomasello et al., 1993). As such, action-copying has been attributed as a central feature of human culture (Boyd & Richerson, 1996; Lewis & Laland, 2012).

One of the most prominent arguments for action-copying as a facilitator of cumulative culture concerns apparent differences in this ability, between humans and non-humans. Call et al. (2005) and Tennie et al. (2009) argued that non-human social learning involves the reproduction of only the *outcomes* of actions ('*product-copying'*), whereas humans can reproduce the actual actions used to reach the outcome ('*process-copying'*). It is this process-copying that is said to boost understanding of what is required to achieve the outcome, thus resulting in the high-fidelity transfer of information. High fidelity information transfer (i.e., the accurate copying of actions responsible for cumulative culture) is argued to be largely responsible for cumulative culture because information that is transferred with accuracy requires much less trial and error on part of the receiver (Tomasello, 1999).

This distinction in action-copying is supported by studies demonstrating that chimpanzees will adopt product-copying in circumstances where causal information about the outcome is clear, whereas human children will copy the actions of a demonstrator regardless of clarity (Horner & Whiten, 2005). Humans of all ages, but not non-humans, are often found to *overimitate*, that is, copying irrelevant actions of others even when they know that they are not necessary to achieve the desired outcome (Lyons et al., 2007; McGuigan et al., 2011). Such behaviour suggests inferences have been made about the intentions of the actions, since copying irrelevant actions does not help to achieve the physical outcome, but may achieve some additional goal regarding the processes involved.

However, process-copying has been demonstrated in many non-human species using the '*do as I do*' paradigm, in which participants are trained to copy the action of an agent, and after training is complete, a novel action is produced. If participants copy the novel action, action-copying is said to have taken place. This has been found to varying degrees in many species including chimpanzees (Custance et al., 1995), orangutans (Call, 2001) and dogs (Topál et al., 2006), suggesting action-copying is not exclusive to humans or primates more generally. Furthermore, there are examples of both human children and chimpanzees successfully learning how to complete a task, but only following social demonstrations, as opposed to a ghost display that shows the process of achievement without a social demonstrator (Hopper et al., 2007, 2010). This suggests that, at least to some degree, non-humans do rely on process-coping, which calls into question its relevance for cumulative culture.

In addition to the evidence that suggests both humans and non-humans are similarly capable of process-copying, there is also growing evidence that this capacity is not necessary for cumulative culture to occur. Caldwell & Millen (2009) found that cumulative culture can occur in circumstances where only access to product-copying was available, and thus process-copying is unlikely to be a (sole, at least) contributor to distinctively human cumulative culture.

Innovation

Even if the capacity to imitate is present, without modification, traits cannot evolve either in terms of refinement or complexity, therefore innovative modifications are necessarily critical to the process of cumulative culture (Enquist et al., 2008). As such, innovation has been suggested as a driver of human unique cumulative culture because it allows an individual to build on the performance of others over and above what can be achieved in populations of non-humans (Vaesen, 2012). While innovation can occur through spontaneous and unintentional production of a novel behavioural variant (e.g., as a result of errors in transmission), those cases most likely to contribute to cumulative culture are often considered to be driven by problem solving, whether it's finding a novel solution to a new problem, or an old problem using behavioural modifications that were not previously available (Kummer & Goodall, 2012; Reader & Laland, 2003).

However, there exists some evidence suggesting innovation is not unique to humans. Anecdotal examples of innovative behaviour in non-human populations include, but are not limited to, birds adapting migratory behaviours to avoid starvation (Sol et al., 2005; Sol et al., 2005), food washing practises in Japanese macaques (Kawai, 1965), and lemurs using their tail to reach and soak up drinking water (Hosey et al., 1997).

Rather than capturing innovation as a novel variant on behaviour, experimental studies focus more on the ability to spontaneously and adaptively switch behaviours to favour a more suitable one following a change in circumstance. This has revealed some propensity for innovation in terms of the flexibility of foraging techniques in great apes (Manrique et al., 2013), lemurs (Dean et al., 2011) and starlings (Boogert et al., 2008).

Therefore, given that innovation is regularly seen in non-humans, there is good reason believe that something else is restricting cumulative culture in non-human populations.

Behavioural Conservatism and Flexibility

Another capacity that has been considered as a driver of cumulative culture is flexible solution switching. Having this capacity means that when a behaviour is not particularly efficient (such as eating soup using a fork – not impossible, but very slow), it can be rejected in favour of a more beneficial alternative (eating soup using a spoon).

While humans are very good at switching solutions, non-humans have been argued to be more conservative, meaning they have difficulty switching solutions even if a better solution is available (Dean et al., 2014). There are numerous reports of great apes retaining a known solution to a problem, even when a novel and more efficient strategy is presented to them (Gruber, 2016; Hopper et al., 2011; Hrubesch et al., 2009; Marshall-Pescini & Whiten, 2008). As such, behavioural conservatism has been suggested to be a barrier to successful cumulative culture in non-human populations.

However, there have been some reports suggesting some non-humans do show a degree of flexibility in their problem solutions, and they do not always perseverate with the responses they have performed previously. For example, Dean et al. (2012) found that chimpanzees continued to explore the features of a puzzle box even after they had developed a solution, suggesting that they were open to the possibility of an alternative. Lehner et al. (2011) found the same exploratory behaviour in captive orangutans, extending to them switching to a more beneficial solution if they found one. As such, it cannot be surely stated that behavioural conservatism is what restricts cumulative culture in non-humans.

In summary, the claims that cumulative culture is enhanced in humans because of an increased potential for imitation and/or innovation, or restricted in non-humans because of behavioural conservatism are not supported fully by empirical evidence. In at least some contexts, non-humans have been found to imitate, innovate and overcome behavioural conservatism, indicating that the absence of cumulative culture in non-human species cannot be attributed to any of these capacities alone. However, there are still seemingly human-unique capacities that have yet to be explored in relation to cumulative culture. It is specifically the cognitive capacities found to emerge in typically developing humans at around 4 years of age, that are particularly promising as a potential explanation for the distinctiveness of human cumulative culture. In this thesis, I explore one of these capacities - mental state understanding - as a possible supporting capacity, that may be responsible for revolutionizing social learning and cumulative culture in humans. In the following section, I discuss the evidence that mental state understanding is unique to humans, and further evidence that it is potentially linked to cumulative cultural evolution.

1.3 The extent to which mental state understanding is unique to humans

1.3.1 The historical context of mental state understanding research

Mental state understanding refers to the ability to attribute mental states to, or infer mental states of others. This includes making reasoned inferences about states such as knowledge, desires and beliefs, in the absence of explicit signals (such as being verbally told they exist) (Premack & Woodruff, 1978a; Spaulding, 2020).

The earliest study to explore mental state understanding investigated chimpanzees' ability to attribute goals to human agents by assessing their ability to identify the tool needed for the experimenter to complete a task (Premack & Woodruff, 1978a). However, the interpretation of this study was controversial, with claims emerging shortly afterwards that chimpanzees' success in this task could be explained more parsimoniously by associative learning (Savage-Rumbaugh et al., 1978). This would mean that goal attribution in these circumstances was not connected to mental state reasoning, and that chimpanzees were forming associations between the objects and the agent's actions, which does not require any understanding of intentionality, beliefs or desires (Gergely & Csibra, 1997).

Since the aforementioned early work, there has been long and extensive debate about whether mental state inference is unique to humans, and in what way. Some argue that mental state understanding is a cognitive mechanism that is not available to non-humans, or human children under 4 years of age, however there is extensive debate about whether this is the case. The controversy surrounding mental state attribution in both non-humans and human children exists mainly because the conclusions drawn from empirical tests depends strongly on the methodology used. These methods are broadly categorised into '*elicited'* and '*spontaneous'* measures. In some of the existing literature, these measures are referred to as '*explicit'* and '*implicit'* respectively, however, because these labels imply that mental state understanding operates at two distinct levels (i.e., just one of the several theoretical viewpoints regarding the contrasting results), the following sections will refer to these measures as '*elicited'* and '*spontaneous'*. These terms accurately distinguish the studies discussed, and are more theoretically neutral than other labels. Elicited measures include those which require a verbal or motor response, such as a point, usually in response to a question which directly addresses understanding. Spontaneous measures, on the other hand, include responses that do not require such a deliberate answer, for example, examination of eye movements. From these (broadly involuntary) responses, inferences are made about the cognitive processing that is likely to have produced them. Only humans over the age of 4 years are reliably found to pass elicited tests of mental state understanding such as false belief reasoning tests (section 1.3.2), however, claims of the ability to understand mental states in younger children and non-humans has emerged from the results of spontaneous tests. The following section will review the methodologies for testing mental state understanding and consider their validity in accurately assessing it. Table 1 summarises some high-profile claims for mental state understanding in child and non-human populations under the methodologies discussed. This table highlights key studies and is not intended to reflect a comprehensive record of the literature.

Table 1

Summary of experimental tests claiming to capture mental state understanding in human and non-human populations.

Traditional False Belief tasks

Methodology	Population	Task Type	Group Tested pass/fail	Age threshold found for passing	Citation	Main finding
Unexpected Transfer	Human	Elicited	3- to 9-year-old children	4 years old	Wimmer & Perner (1983)	An age-related increase was found in the ability to correctly identify location of search based on false belief with age.
			3.5- to 6-year-old children	4 years old	Baron-cohen et al. (1985)	Typically developing, but not autistic children were able to correctly identify the location that the agent would search if they had a false belief.
			4- to 5-year-old children	4 years old	Call & Tomasello, (1999)	Scores in a nonverbal false belief task correlated with a verbal false belief task – children at 4 years were found to reliably pass both.
			4- to 5-year-old children	4 years old	Krachun et al. (2009)	Children responded correctly in both true and false belief tasks by 5 years.

Spontaneous, violation of expectation	15-month-old children	15 months old	Onishi & Baillargeon (2005)	Infants looked for longer at a scene when the agent's actions did not match the child's knowledge of the agent's belief.
	13-month-old children	13 months old	Surian et al. (2007)	Infants looked for longer at a scene when the agent's actions did not match the child's knowledge of the agent's belief.
	18-month-old children	18 months old	Song et al. (2008)	Infants looked for longer at a scene when the agent's actions did not match the child's knowledge of the agent's belief.
	7-month-old children Adult humans	7 months old	Kovacs et al. (2010)	Replication study of Onishi and Baillargeon: Both 7-month-olds and adults looked longer at the scene when the agent's belief (true or false) was not confirmed by their actions.
	15-month-old children	15 months old	Träuble et al. (2010)	Children looked for longer at an agent's reach that is not consistent with the agent's goal
	14-month-old children	14 months old	Poulin-Dubois et al. (2013)	Infants took the agent's perspective into account even if it did not match with their own perspective.
	14- and 18-month-old children		Yott & Poulin-Dubois (2016)	Neither age group indicated false belief understanding by looking for longer at a scene when the agent's actions did not match the child's knowledge of the agent's belief.

				No inter-task correlation found between belief and desire.
	18-month-old children		Powell et al. (2018)	Found no evidence that 18-month-old' children's expectations about an agent's beliefs were violated when the agent's actions did not match the child's knowledge of the agent's belief.
	18-month old children		Poulin-Dubois & Yott (2017)	No indication of false belief understanding by looking longer at a scene when the agent's actions did not match the child's knowledge of the agent's belief. No inter-task correlation found between tasks replicating (Buttelmann et al., 2009; Onishi & Baillargeon, 2005).
Spontaneous, Anticipatory looking	2- to 4-year-old children	2 to 4 years old	Garnham & Ruffman (2001)	Anticipatory looks to correct locations indicated use of implicit mental state understanding rather than alternative explanations (associative learning/seeing = knowing).
	25-month-old children	25 months old	Southgate et al. (2007)	Children correctly looked in anticipation to the location where the agent will look if they have a false belief.
	Neurotypical human adults and human adults with		Senju et al. (2009)	Neurotypical adults, but not adults with Asperger's syndrome, spontaneously anticipated the action of an agent, signalled by anticipatory looking.

	Asperger's syndrome			
	Adult humans		Schneider et al. (2012)	Adult's anticipatory looking suggested that they follow the beliefs of the agent without being explicitly told to do so.
	11- to 17-month-old children	17 months old	Surian & Geraci (2012)	17-month-olds but not 11-month-olds anticipated the actions of a shape agent (triangle) using anticipatory looking.
	3- to 4-year-old children Adult humans	3 years old	Low & Watts (2013)	All age groups correctly anticipated the actions of others, signalled by anticipatory looking. Only 4-year-olds and adults were able to make verbal predictions about the agent's actions.
	Adult humans		Kulke et al. (2018)	Failed to replicate using the methodology of anticipatory looking results of the following studies: (Low & Watts, 2013; Schneider et al., 2012; Senju et al., 2009; Southgate et al., 2007; Surian & Geraci, 2012)
	24-month-old children		Dörrenberg et al. (2018)	Failure to replicate direct test of second-order false beliefs, but did replicate both first-order and true beliefs both
	2- to 8-year-old children Adults		Burnside et al. (2018)	No age group demonstrated a clear understanding of false beliefs through anticipatory looking.

		4.5- to 5.5 year old children Adults	Kulke et al. (2019)	No anticipatory looking behaviours found, except for in second-order false belief understanding in adults. Inconsistent patterns of belief tracking in usually converging studies (lack of convergent validity).
Non-humans	Elicited	Chimpanzee Orangutan	Call & Tomasello (1999)	Apes selected locations by touching. None passed the false belief task.
		Chimpanzee Bonobo	Krachun et al. (2009)	Apes always reached for the location that the agent reached for – so they were correct in true beliefs, but incorrect in false beliefs. Face orientation indicated some possibly implicit cues, but this was very fragile.
	Spontaneous, anticipatory looking	Chimpanzee Bonobo Orangutan	Krupenye et al. (2016)	All apes looked in anticipation to the location where they expected the agent to search for an object in accordance with false beliefs.
		Chimpanzee Bonobo Orangutan	Kano et al. (2019)	All apes subscribed a false belief if they witnessed the act that caused the false belief to happen.
		Japanese Macaques	Hayashi et al. (2020)	Macaques anticipated false-belief guided actions using anticipatory looks.

			Rhesus Macaques		Marticorena et al. (2011)	Looking times were no different in expected and unexpected events. Failed to replicate the Onishi and Baillargeon paradigm in non-humans.
			Rhesus Macaques		Martin & Santos (2014)	Macaques looked for longer at events that violated their own expectations about the location of the object, but their looking time was not affected by an agent's belief, suggesting they did not spontaneously track other's beliefs.
Unexpected Contents	Human	Elicited	3- to 4-year-old children	4 years old	Perner et al. (1987)	Children correctly attributed a false belief to an agent at 4 years of age, but younger age groups had difficulty with this.
			3- to 5-year-old children	4 years old	Gopnik & Astington (1988)	Children were able to consider representations from another's perspective based on their false belief from 4 years of age.
		Spontaneous	3.5- to 5-year-old children	4 years old	Krachun et al. (2010)	Developmental progression was found for false belief understanding in children (performance was below chance at 3.5 years of age, but above chance at 4.5 years of age).
			Chimpanzee			Chimpanzees spontaneous looking behaviours showed no evidence that they were capable of recognising the false beliefs of the agent.

Goal Attribution tasks

Methodology	Population	Task Type	Group Tested pass/fail	Age threshold for passing	Citation	Main finding
Goal Attribution	Human Children	Spontaneous, unexpected transfer, active helping	18-month-old children	16 to 18 months old	Buttelmann et al. (2009)	By 18 months, children were found to consider an agent's belief when interpreting their goal. Also found some evidence of this capacity emerging at 16-months.
		Spontaneous, violation of expectation	1-year-old children		Gergely et al. (1995)	Children looked for longer at a scene when the agent's action is irrational based on their goal.
	Non-human	Spontaneous, active helping	Chimpanzee		Premack & Woodruff (1978b)	Chimpanzees were found to reach for objects required by an actor when those objects were out of reach to the actor.
			Chimpanzee Bonobo Orangutan		Buttelmann et al. (2017)	Used the same experimental paradigm as Buttelmann et al. (2009) with non-human primates. Found that apes behaved in the same way towards an agent's false belief as a human would.
		Spontaneous, violation of expectation	Chimpanzee		Uller & Nichols (2000)	Used the same methodology as Gergely et al. (1995) to demonstrate that chimpanzees looked for longer at a scene when the agent's action is irrational based on their perceived goal.

1.3.2 Explicit measures identifying the capacity for mental state understanding in children from 4 years of age

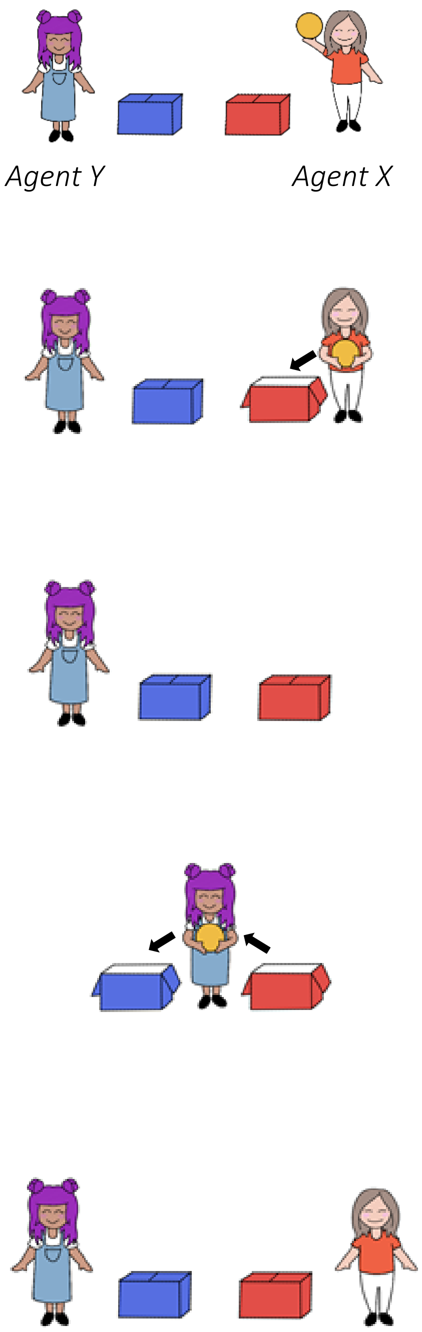
False belief attribution tasks

Elicited response tasks are direct measures, in that they require a response to a direct understanding-based question. Elicited tests of false beliefs typically set the age of mental state attribution at around 4 years of age, with children under this threshold reliably found to respond incorrectly to test questions.

The gold standard test of mental state understanding is widely regarded to be the *false belief attribution paradigm*. These tasks require a participant to make an inference about an agent's belief that is different from their own. False belief attribution tasks were initially conceptualized by Dennett (1978), and first implemented using child participants by Wimmer & Perner (1983) in the *'unexpected transfer task'* (Figure 1). In this task, a child is shown a scene with two characters (for all similar examples in this chapter, these characters will be referred to as 'agent *X*' and 'agent *Y)*'. Agent *X* hides an object in location A before leaving the scene. In the absence of the agent *X*, agent *Y* displaces the hidden object to location B. When the agent *X* returns, the child is asked *'where will agent X look for the object'*. The child can either verbally express or point to a location to give their answer. Here, children are being asked to make an inference about the agent *X*'s belief about the object's location, and if they have a capacity for mental state understanding, they should therefore correctly infer that the agent *X* has a false belief by stating that they believe the object to be in location A. Both Wimmer & Perner (1983) and Baron-cohen et al. (1985) found that neurotypical children become capable of correctly answering only at around the age of 4 years (Of course, it is worth noting that there is some individual variation with this, as some children pass the test as early as 3 years and others as old as 5 years). Prior to acquiring this ability, children typically respond (incorrectly) with the actual location of the target object (location B). Therefore, the younger children appear to fail to understand that agent *X* should falsely believe that the object is still in the location where they had put it.

Figure 1

The traditional false belief task, adapted from Wimmer & Perner (1983)



This is Agent *X* and Agent *Y*.

Agent *Y* has a blue box.

Agent *X* has a red box.

Agent *X* has a marble.

Agent *X* puts the marble in her red box.

Agent *X* goes for a walk.

Agent *Y* takes the marble out of the red box and puts it in the blue box.

Now Agent *X* has come back.

Where will Agent *X* look for their marble?

A related paradigm, the '*unexpected contents task'* (Gopnik & Astington, 1988; Perner et al., 1987) demonstrates that children under 4 years of age also fail to attribute false beliefs to themselves. In this paradigm, the child is given an opaque container labelled to contain sweets. When the child is asked what is inside the container, they correctly (according to the only piece of evidence available) answer *"sweets"*. The container is then opened to reveal its actual contents – some pencils. When asked what their friend will think is inside, children under 4 years of age typically answer egocentrically, claiming their naïve friend will share their knowledge that what's inside is pencils. Interestingly, when knowledgeable children are asked to recall what they originally thought was in the tube, they also answer "*pencils*". Therefore, not only are they unable to attach a false belief to someone else, but they are also unable to attach a false belief to their past self. This is in line with research reporting that children will claim they have always held knowledge that they had just learned, which holds true until around the age of 5 years (Taylor et al., 1994).

Due to the requirement of classic tests of elicited false beliefs to include verbal communication with the participant, such studies have not been implemented in infants, or in non-human primates. Given that no conclusions can be drawn about these populations in classic elicited response tests, simplified spontaneous response measures have been used to assess the presence of mental state understanding in these populations.

While some argue that elicited false belief tasks are the only way to demonstrate mental state understanding (Gomez, 2004), others argue that an earlier developing mental state understanding mechanism in children under 4 years of age might indicate an earlier development of mental state understanding, measured in a way that is not affected by inhibitory control (Leslie, 2005) or limited executive function (Rubio-Fernández & Geurts, 2013). The following section discusses tests using spontaneous response measures of mental state understanding in non-humans and children under 4 years of age.

1.3.3 Spontaneous measures identifying the capacity for mental state understanding in non-humans and children under 4 years of age

Evidence for mental state understanding in young children (under 4 years of age) emerges from the same methodology that claims to find evidence in non-humans. Of course, viable evidence of mental state understanding in both populations would have considerable implications for our understanding of cumulative culture. Demonstrating the use of mental state understanding in children under the threshold 4 years could suggest that adult-like mental state understanding is merely enhanced, rather than enabled by later developing mechanisms, which could bring into question its seemingly important role in cumulative culture. Similarly, a presence of mental state understanding in non-human populations would call into question claims that it is a distinctively human mechanism, and its role in cumulative culture would be unlikely to be what sets it apart from the capacity in humans. However, such claims are based on controversial spontaneous response measures, which, unlike the directness of elicited measures, rely on inference about participant understanding on the part of the experimenter. As such, spontaneous response measures are more open to interpretation than elicited response measures, and it is entirely possible that the results of these tasks pick up false positives that may have come about for reasons other than the participant reasoning about mental states. Furthermore, it could be argued that even if these tasks really do capture mental state understanding, this may be unlikely to be beneficial to cumulative culture relative to an understanding which can be subjected to explicit reasoning about others' knowledge. The following section assesses the spontaneous response measures used to test mental state understanding in non-humans and children under 4 years of age.

1.3.4 Evidence of mental state understanding in non-humans and very young children

Tactical deception

Non-human mental state understanding is commonly associated with tactical deception tasks, which typically demonstrate a non-human primate acting in a certain way that appears to cause a false belief in another. On the surface, tactical deception creates a convincing basis for arguing that mental state understanding is present in non-humans – after all, to deceive, the deceiver has to be aware that another agent can have a mind state that differs from their own, and an understanding of how they can behave to manipulate that mind state. However it is also possible, and highly likely, that the behaviour seen in non-humans in these tasks is resting on mechanisms other than mental state understanding.

There is substantially more evidence of tactical intent to deceive in non-human primates relative to human children (Byrne & Whiten, 1992). Non-human primates are seen to deceive by withholding the location of food (Hirata & Matsuzawa, 2001; Melis et al., 2006), producing gaze and movement cues to manipulate others to attend sites of low reward to free up space at a site of high reward (Hall et al., 2017) and by using alarm calls to distract other's from acquiring food (Wheeler, 2009)). Despite this wealth of compelling evidence that non-human primates may be capable of mind reading with intent to deceive, their behavior could be explained as having occurred by chance, or by use of alternative mechanisms such as trial-and-error learning based on previously experienced similar situations (Heyes, 1998).

Measuring eye gaze

False belief tasks have been adapted to explore potentially successful attribution of mental states in non-humans, and children under the age of 4 years. These adapted tasks typically measure spontaneous cues from eye movements to infer children's expectations about where an agent will search for a hidden object. Eye movement recording therefore expands the false belief paradigm by allowing assessment of non-verbal participants, and assessment of involuntary responses.

Anticipatory looking in unexpected transfer tasks has been used to study children too young to pass the tests which use elicited responses. This spontaneous understanding is assumed based on the child's direction of gaze during the task, such that gaze towards a location indicates that the child expects some activity to occur at that location. Looking in anticipation has been demonstrated in 2-year-old children, in the absence of any explicit response (Garnham & Ruffman, 2001; Southgate et al., 2007), leading to claims that children are able to attribute false beliefs to agents much earlier than they are able to explicitly demonstrate this capacity.

Anticipatory looking tasks are primarily used because they can be easily adapted for use with non-humans, while preserving the basic logic of the unexpected transfer task (Kano et al., 2017). One of the most recent studies to be carried out in non-humans, implemented by Krupenye et al. (2016), used a scene layout adapted from the traditional unexpected transfer task, whereby an agent hides an object, only to have it displaced by another agent in their absence. When the hider comes back, the participant's direction of gaze gives a spontaneous cue to where they expect the hider will search for their hidden object. In this task, apes, including chimpanzees, bonobos and orangutans, were found to reliably look at the location the hider originally placed the object, rather than the location it was last seen by the participant. This suggests that apes may have some understanding of the protagonist's belief about the object location. More recently, the same paradigm has been used to demonstrate similar performance in Japanese macaque monkeys (Hayashi et al., 2020).

One of the most convincing studies of mental state understanding in children under 4 years of age was conducted by Clements & Perner (1994). The main difference between this study and other spontaneous tests is that the child's response (a look in anticipation) was made following an explicit prompt from the experimenter (e.g., "*I wonder where he's going to look?*"). This means that the child's response is open to much less interpretation than other studies, which rely on children following the story on their own. Interestingly, this study only found anticipatory looking from children at around 3 years of age, preceding elicited mental state understanding much more closely than other tests requiring a spontaneous response. This calls into question the validity of similar tests that find positive results for anticipatory looking from much younger children.

Looking times

Evaluation of gaze to assess reasoning about mental states is not limited to anticipatory gaze. It is also possible to measure looking time, for example, the length of time a participant spends attending to a scene, or feature of a scene. This measure requires a similar task structure to the anticipatory looking tasks, however, instead of assessing the direction of gaze before an event, it measures the duration of gaze following an event. A longer duration of gaze indicates that something has occurred that does not match the participant's expectations, and a shorter gaze indicates dismissal due to the participant's expectations matching the outcome of the event. These tasks have led to claims that very young children, from the age of 3 months are capable of detecting events that are unexpected (Spelke et al., 1992).

Looking times are therefore used to assess an infant's reaction to unexpected events, with longer looking times indicating surprise (i.e., a violation of expectation) at an event that did not have the outcome that matched to the expected outcome based on earlier activity in the scene. Holding an expectation about an agent's behaviours is assumed to require some level of mental state understanding to represent an agent's goal, and to distinguish behaviours that are consistent with their goal (i.e., rational actions) from those that are not consistent with their goal (i.e., irrational, unexpected actions). Gergely et al. (1995) illustrated the '*violation of expectation paradigm*' by habituating participants to a visual scene of a ball character moving from the left side of the screen to the right by jumping over a visible obstacle (the jump being a rational action), or jumping over an invisible obstacle (the jump being an irrational action in the absence of the obstacle). Participants were then shown two test events with no obstacle – one where the ball moved to the other side of the scene along the floor without jumping, and another where the ball moved to the other side of the screen by jumping over an invisible obstacle. Participants who observed the ball character being rational in the habituation phase were expected to predict that the ball character will move to the other side of the screen in the most rational way in the test event, evidenced by a reduced looking time. This was the case – children at around 12 months of age spent much longer looking at scenes where the ball character made an irrational jump compared with scenes when the ball character moved without jumping. This suggests that these children had some expectation about the character's goal-directed behaviour based on their behaviour in the habituation event, as well as the agent's intention – which is only to move to the right side of the scene. Using an adapted version of the same task, Uller & Nichols (2000) found the same result as in the chimpanzees, suggesting that they were also able to attribute the correct goal to the moving agent. While goal attribution may be a feature of mental state understanding, however, it is not clear from these tasks whether infants or chimpanzees could recognise circumstances where the agent has a different goal, or belief, to their own.

Violation of expectation has been assessed in a similar context to the traditional unexpected transfer task, revealing false belief attribution in children as young as 15 months of age (Onishi & Baillargeon, 2005), and distinction between self and other beliefs in children aged 13 months (Surian et al., 2007). Furthermore, using this task design, children aged between 14 and 18 months have been claimed to be able to update their knowledge of other's beliefs, for example, Song et al. (2008) found that when the protagonist of the unexpected transfer scene was given information about the correct location of their object following its displacement, 18-month-old children looked for longer at the event when the agent searched in the empty location, suggesting the children had updated their understanding of the protagonist's knowledge. To check that this effect was not related to the child's own visual access (i.e., updating their own knowledge rather than the protagonist's), Poulin-Dubois et al. (2013) used an adapted version of the unexpected transfer task used by Onishi & Baillargeon (2005), in which the protagonist was blindfolded during the search. With regards to looking times, children showed no expectations about where the protagonist would search, suggesting that infants were disregarding their own visual access to take the protagonist's ignorance into account.

Active helping

Despite the controversy surrounding goal attribution, a compelling contribution to the mental state understanding literature has been made by use of the active helping paradigm, the basic principles of which involve a participant appropriately assisting an agent with a potentially differing goal. Supporters of this paradigm argue that active behavioural response to the unexpected transfer task carries more weight than visual response measures. For example, Buttelmann et al. (2009) claimed that visual response tasks only indicate that the child recognises something is unusual, rather than identifying what it is they find unusual, meaning that it is possible that these tasks do not require an understanding of mental states. In an active helping task, Buttelmann et al. (2009) found that by 18 months, children will attempt to assist an agent with their goal, in ways that depend on the belief the agent holds about the location of an object in the unexpected transfer task. To successfully assist the agent, the child must have some understanding of the agent's goal, and belief about the location of the object. For example, if the agent holds a false belief and searches location A (i.e., the original location of the object before displacement), the child will help them look for the toy in location B (i.e., the true location of the object after displacement). However, if the agent holds a true belief that the object is in location B, but they search location A, the child will help to search in location A, assuming that their true belief about the object being in location B means that their goal is not to locate the object. In a later study, Buttelmann et al. (2017) found that great apes (including chimpanzees, bonobos and orangutans) also performed similarly to children in these tasks, apparently successfully applying an understanding of false beliefs.

1.3.5 Failures to replicate classic spontaneous tests

Recent evidence has emerged suggesting that many of the tasks using spontaneous measures do not produce reliable results. Many researchers have now attempted to replicate previous studies, failing to find the same results, sometimes using larger samples. Furthermore, different measures, such as anticipatory looking and violation of expectation, do not appear to reliably capture the same results at the same ages, indicating a lack of convergent validity. Failures to reproduce key results have been found in direct replications of both human infant (Kulke et al., 2019; Kulke & Rakoczy, 2018) and non-human primate (Horschler et al., 2020) studies, and conceptual replications (close, but not exact replications) (Dörrenberg et al., 2018; Poulin-Dubois & Yott, 2017; Powell et al., 2018; Yott & Poulin-Dubois, 2016). Furthermore, a large number of unpublished studies have been collated by Kulke & Rakoczy (2018), many stating a failure to replicate using both direct and indirect measures.

1.4. Alternative explanations for the results found in implicit tests

The previous section outlined some of the experimental claims for spontaneous mental state understanding in non-humans and in young children, however these claims are based on methodologies that rely on *inference* about behaviours rather than explicit evidence that these behaviours are caused by an understanding of other's mental states. As such, these tests are not without controversy, even setting aside the aforementioned replication problems. The following section will discuss some of the dissenting voices contributing to the wide variety of alternative, non-mentalistic explanations for the positive results found in these populations.

In a recent article, Quesque & Rossetti (2020) stated two main criteria that should be met by measures of mental state understanding. Firstly, it should be clear that respondents maintain a distinction between their own and other's mind states. Secondly, success on a task cannot be accounted for by lower level processes such as arbitrary attention orientation and associative learning. However, it is entirely possible that all of the examples of spontaneous mental state understanding discussed above can be explained using simpler alternative mechanisms that do not require mental state inference.

One of these mechanisms, proposed by Povinelli & Vonk (2004), details a parsimonious system whereby inferences about others are derived from observable behaviours alone. This system states that social interaction only very occasionally involves generating inferences about mental states. The vast majority of inferences are derived from reading the *behaviours* of others, and associations between these behaviour and the environment (i.e., associative learning). Associative learning, which is generally found to be ubiquitous in all animal species, can account for the majority of behaviours usually attributed to mindreading (Penn & Povinelli, 2007; Heyes, 2012). Heyes (1998) stated that in every study claiming non-human primates have an understanding of mental states, the behaviour could have occurred by associative learning or other non-mentalistic capacities such as behaviour reading. Infrequent mental state inference (i.e., in rare situations where reading behaviour is not sufficient) may occur in human adults, but it is likely that human infants and perhaps non-humans rely on behaviour reading mechanisms even in cases where it may appear that they are applying an understanding of mental states.

Others argue that the results found in tests of spontaneous mental state understanding capture the '*teleological stance*' (Gergely & Csibra, 1997, 2003). Teleology refers to action interpretation based on reality alone. This is done using '*the principle of rational action*', which assumes that the agent will act only to fulfil their goal by the most efficient way possible. This is thought to involve the representation of actions using three relevant situational aspects, each of which represents a non-mentalistic concept: the agent's action, the agent's goal, and environmental constraints. Infants may be able to make a non-mentalistic inference about one of the three situational aspects if they have access to information about the other two, for example, in the violation of expectation task, knowing the goal of the agent and the environmental constraints, infants are able to predict an efficient action.

While teleology does, in theory, stand as a potential mechanism for predicting actions, goals or environmental constraints in children with immature mental state understanding, a lack of direct empirical support for these claims means that its use as an alternative explanation must be treated with some caution (Juvrud & Gredebäck, 2020). Furthermore, it has been proposed that any experiment effective enough to test for mental state understanding independently from behaviour reading, associative learning or teleology (i.e., elicited response tests) would not be suitable for use with any non-human or human infant due to the cognitive demands necessary to glean the desired response (van der Vaart & Hemelrijk, 2014; Buckner, 2014).

One of the highest-profile alternative explanations is the '*two-systems theory*', which suggests that children and non-humans passing spontaneous-response tests of mental state understanding are doing so using an entirely different system for inferring other minds than human adults. Apperly & Butterfill (2009) proposed that mental state understanding develops in at least two distinct systems, one of which develops earlier and elicits the automatic but limited response seen in young children and non-humans (system-one), and the other later developing system, which allows access to more flexible processes that neither of these populations have (system-two). The crucial difference between these systems is in their representational capacity. For example, while a young child using system-one can recognise *whether* an agent can see an object, only children using system-two can form a representation of *how* an agent can see an object, including distinguishing when the agent's representation differs from their own.

Experimental examinations have found evidence of a separate automatic system for perspective taking in adults (Furlanetto et al., 2016; Qureshi et al., 2010; Samson et al., 2010), although see Cole et al. (2016) and Conway et al. (2017) for an alternative explanation). However, the results of these studies have been attributed to '*submentalizing'* (Heyes, 2014), which claims participants in these tasks use behavioural cues rather than applying mental state understanding. This theory has been supported by evidence suggesting that the same automatic perspective taking effect is found when character agents are replaced with arrows (Santiesteban et al., 2014). As it stands, there is no conclusive evidence that adult humans have access to system-one, and therefore, there is limited support for the two-systems theory. Of course, as noted previously, even if there is an early, but limited, capacity for mental state understanding (as would be consistent with the two-systems theory), it is unlikely that the level of understanding achievable using simpler systems would be sufficient to underlie the behaviours required to facilitate cumulative culture.

The current section has outlined some of the alternative explanations proposed for the positive results found for mental state understanding tests in non-humans and very young children. Alongside the failures to replicate studies testing for mental state understanding using spontaneous measures, alternative non-mentalistic explanations for positive results add weight to the idea that reasoning about mental states (which may be critical for cumulative culture, as we will argue below) is not present in non-humans, nor very young children. In contrast, tests of mental state understanding that use elicited responses are relatively uncontroversial, and remain consistent with theories claiming that 4 years of age is a key milestone in cognitive development. The following section will discuss further evidence in support of these claims.

1.5 Further evidence of age 4 as a key cognitive milestone

Meta-representational flexibility

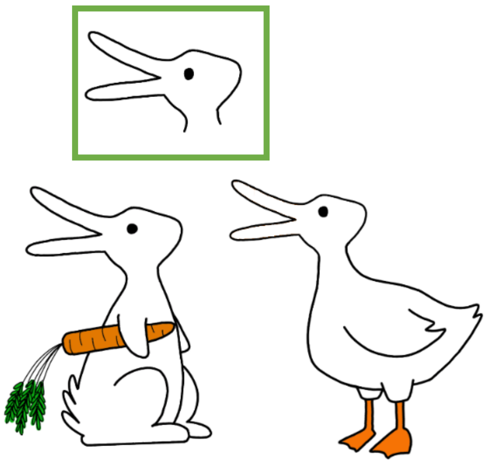
Given the wealth of evidence suggesting that tests of spontaneous mental state understanding may not be capturing the same capacity in young children and non-humans that is seen in children over 4 years of age and adults, it appears that this age represents a key milestone in distinctively human cognitive development. Tests of meta-representational flexibility further support the significance of this developmental milestone, with evidence suggesting that older children are able to adopt a kind of flexibility of reasoning that is not available to younger children. This adds weight to the idea that the age at which children pass elicited response false belief tasks reflects a real change in their conceptual understanding.

Tests of meta-representational understanding typically measure the ability to hold multiple representations of the same object/person, and the ability to flexibly switch between these representations. It makes logical sense that holding multiple representations of the same person or object is a fundamental requirement for passing the unexpected transfer test (which requires a representation of both reality and the agent's false belief). However, to understand that someone else may hold a different number of representations than oneself, one must first have an understanding that an object *can* be represented in several ways.

Children under 4 years of age typically struggle to recognise more than one label for an object. Supporting evidence of children's difficulty with holding multiple representations is found in the '*ambiguous figures*' test (Doherty & Wimmer, 2005). An ambiguous figure is a picture that can be fully interpreted in more than way. For example, the shape in Figure 2 can be viewed as either a duck or a rabbit. While it's possible for adults and older children to freely switch from rabbit to duck and back again, for children below the age of acquisition for mental state understanding, this is problematic even in circumstances where they are aware of both possible interpretations. This demonstrates that children under 4 years of age are unable to represent an object as two things at once. Acquisition of this meta-representational understanding indicates the emergence of higher-level cognitive skills that support human-unique culture.

Figure 2

Recreated stimuli from the ambiguous figures test, adapted from Doherty & Wimmer (2005)



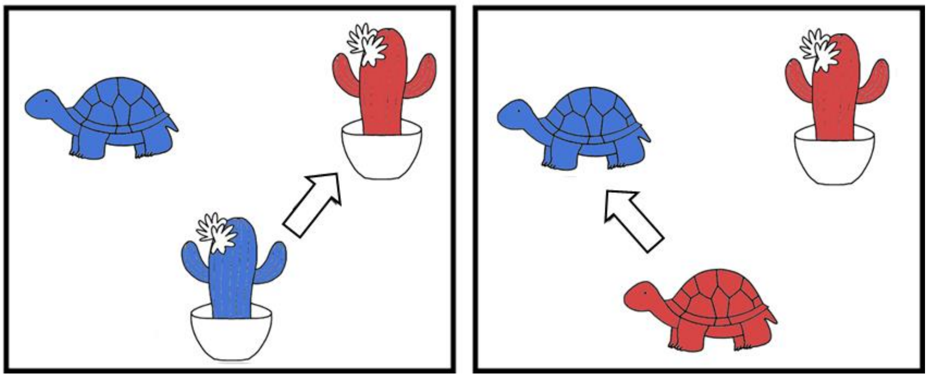
Note. Participants are shown the image in the green box, which could be seen as a rabbit's face or a duck's face. Images of the full duck and rabbit are included here to demonstrate how the ambiguous figure could fit both interpretations.

The cognitive turning point at 4 years of age is also measured using tests of more general attentional flexibility and cognitive control, for example, in the ability to see the same object as belonging to more than one dimension, and flexibly switching between them. This capacity is evaluated using the dimensional change card sort task (Figure 3; Zelazo et al., 1996), which generally finds that children are able to do this effectively by

5 years of age. In this task, children are given test cards depicting coloured objects (such as '*blue tortoise*', and '*red cactus*') and categories in which to sort those cards (such as boxes labelled with a picture of a blue cactus and red tortoise). First, the child is asked to sort the cards by one dimension, for example, by type, so the cactus goes in the cactus box and the tortoise in the tortoise box regardless of colour. Afterwards, children are asked to switch dimension, and sort by colour. Now, the cactus belongs in the tortoise box and the tortoise belongs in the cactus box. While 5-year-olds and the majority of 4-year-olds are able to do this without any difficulty, 3-year-olds perform poorly, unable to switch dimensions after the initial sorting, indicating perseveration with the original sorting method, despite being able to correctly state the new rule. It should be noted that the difficulty children have is specifically with the switching element, not with any particular sorting method (Zelazo, 2006). This contributes to the argument that children's general flexibility of perspective is immature before the age of 4 years.

Figure 3

Recreated stimuli from the dimension change card sort task, adapted from Zelazo et al. (1996)



Note. The participant is asked to sort by one dimension (in this case, type, indicated by the arrows), and then will be asked to sort by the other dimension (colour).

1.6 Development of mental state understanding and related capacities beyond 4 years of age

While it is widely accepted that the 4-year milestone carries a cognitive turning point in human development, there is evidence to suggest that the capacities emerging at this age continue to develop much later into childhood. As such, children who have reached the 4-year milestone may have access to basic forms of adult-like capacities such as '*first-order*' mental state understanding, but they may not have access to *higher* orders of mental state understanding until they are older (Doherty, 2008). These higher orders allow more complex reasoning about beliefs. The first-order belief reasoning captured in the standard false belief task allows reasoning of scenarios like the following, against real circumstances:

"Liz believes that the sock is in the green drawer".

As such, a child can represent Liz's belief about the sock's location even if the sock is in the red drawer. Higher-order reasoning allows additional layers of belief representation to be added to this, for example, '*second-order*' belief reasoning would allow an understanding of the following:

"Mark believes that Liz believes that the sock is in the green drawer".

Testing for second order mental state understanding has proved challenging, particularly in terms of capturing age differences in adult samples (Oesch & Dunbar, 2017). In an early study using an adapted traditional false belief task, Perner & Wimmer (1985) found that second-order false belief understanding becomes accessible to children at around the age of 7 years (although see Sullivan et al. (1994) for a simplified version capturing this capacity in 6-year-olds). Even more recent studies have used a similar paradigm (involving 3 agents in the standard false belief task) to capture a *third* order, revealing that even more complex reasoning emerges between 14 and 20 years (Valle et al., 2015). Of course, as the anecdotes used in these tests become more complex, it bears consideration of whether a lack of evidence at certain ages is caused by the language in the task materials being syntactically complex.

Further studies attempting to capture mental state understanding in older children and adults include measures that test the ability to reason about abstract behaviours (strange stories test; Happé (1994)), or taking other's perspectives (Newcombe & Huttenlocher, 1992; Piaget & Inhelder, 1956). An early study reported a continual development of second-order perspective-taking up until the age of 8 years (Salatas & Flavell, 1976), however more recent research revealed further refinement (i.e., making fewer errors) up to the age of 27 years (Dumontheil et al., 2010).

The evidence for further developing mental state understanding into adolescence suggests that there may be further capacities that rely on this more complex reasoning coming into fruition at around the same time. This means that 4 years of age is not necessarily the point at which humans become capable of adult-like reasoning, including the level that is required for cumulative culture. However, it is likely that the age of 4 years is the point at which humans start to develop these distinctive mechanisms. The following section will review evidence that suggests mental state understanding, more generally, is a key prerequisite for cumulative culture.

1.7 Mental state understanding as a key capacity for cumulative cultural evolution

The previous section established that the only convincing evidence of mental state understanding is in humans – specifically those over the age of 4 years. The cognitive turning point at 4 years of age is supported by a wealth of claims stating that not only are there alternative explanations for earlier developing mental state understanding, but that the tests showing an early mental state understanding fail to replicate. This is critical from the perspective of investigating mental state understanding as a key factor facilitating distinctively human cumulative culture, because claims of early development are based on evidence analogous to that used to make claims of mental state understanding in non-human primates. This would therefore bring to light the idea that it could be mental state understanding that distinguishes human from non-human culture. However, arguments for the uniqueness of human capacities for mental state understanding are not the only reasons to believe that it could be implicated in cumulative culture. There are also logical arguments that can be made for why and how capacities for mental state understanding might fundamentally alter cultural transmission, in such a way that beneficial modifications are more likely to be preserved, further transmitted, and modified further in advantageous ways. This section therefore explains how mental state understanding, specifically the distinctive mechanism found in humans over the age of 4 years, could be a key cognitive prerequisite for cumulative culture.

Flexible use of social learning strategies

Having access to mental state understanding allows application of reasoned understanding about others and why they make, or do not make, good models for learning. Of course, exhibiting preferences for good models for learning does not necessarily require mental state understanding - there are many examples of non-humans using '*social learning strategies*' to achieve the same result in some circumstances (Laland, 2004). Social learning strategies are rules that guide who to learn from, such as '*copy when uncertain*', '*copy the most experienced*', or '*copy the majority*'.

Heyes & Pearce (2015) argued that all non-human social learning, and likely a great deal of human social learning, is dependent on general-purpose associative learning. Associative learning processes are not driven by understanding, but rather heuristic biases (that is, forming a positive association that favours copying a successful model as a result of their connection with reward, e.g., food). However, there are instances in which humans use '*explicitly metacognitive'* social learning strategies, which are reportable rules that consciously represent properties of the user's or other's mental states and other cognitive processes. Recent support has emerged for a developmental trajectory of explicitly metacognitive social learning. Blakey et al. (under review) tested children between the age of 4 and 8 years using a task in which children had to select one of two agents to give them information about a scene that only one of the agents had visual access to. It was found that older children selected a knowledgeable informant (i.e., the agent who had visual access to the scene) from whom to learn more often than younger children, suggesting that the older children were likely to have used some reasoning about perceptual access, and thus the knowledge held by the agents.

It is thus proposed that humans' distinctive propensity for flexible, selective social learning drives cumulative culture. That is, selectively using social learning strategies when they are most required, such as in cases when beneficial outcomes are opaque (such as manufacturing processes for safe food preparation (Henrich & Broesch, 2011)), and coupling this with asocial interaction with the environment (Galef, 1995). Therefore, having an understanding of mental states allows humans to make accurate representations of what they themselves know, and also what others know, in order to make an informed decision about when to seek information from a knowledgeable model (including making decisions about what to transmit about their own knowledge, see below), and when to rely on trial and error learning, or innovation (Heyes, 2016). This method is much more effective in the context of cumulative culture compared with biases that are driven by associative learning.

Flexible teaching

Mental state understanding is not just a useful tool used to read more from behaviour of others, but also to tailor signals to the receiver of information. This is an essential feature of intentional teaching, which requires information to be transmitted from a knowledgeable source to a naive source. Even without restrictions on time, it would be impossible for a cultural parent to transmit every detail of the knowledge they have to the cultural offspring, and therefore it is beneficial for the cultural parent to transmit the information to the cultural offspring in a way that inspires extended, general-purpose learning , that, in such a way that helps them to fill in the gaps. In this sense, teaching can form the basis for accumulation over time.

The evidence of teaching in non-humans is restricted only to instances of information transmission whereby a cultural parent actively supports the transmission of information that will promote enduring traits in a cultural offspring (Hoppitt et al., 2008). However, this is distinctly different from *intentional teaching* because, while the cultural parent is producing information that modifies the cultural offspring's behaviour, there is no requirement for the cultural parent to understand that their information transmission is the reason for the resulting behavioral modification. This kind of teaching is referred to as '*functional teaching*', because while the animal is producing information intentionally, it is likely due to an evolved adaption that functions to facilitate learning in others, rather than being caused by a deeper understanding of the learner's needs (as is required in intentional teaching). While transmission of inadvertent information can promote enduring traits through local enhancement (i.e., searching a foraging site because you see someone else searching there (Heyes et al., 2000)) many species are not found to intentionally transmit information to benefit another's uptake of that information (Whiten et al., 2003).

In uncommon examples, functional teaching can be applied to some cases in non-humans (Thornton & Raihani, 2008). For example, Thornton & MaCuliffe (2006) found that wild meerkats will teach their pups to disarm prey by exposing them to scorpions that have had their sting removed, and gradually introducing scorpions with varying levels of disability. Further evidence from Franks & Richardson (2006) revealed existence of '*tandem running'* (i.e., guiding another agent to a specific location, or through a route in intervals to track their progress) in ants. Together these studies suggest that some group-living species have evolved to actively transmit information to facilitate other's learning. These examples of non-human teaching are heavily restricted by context, even though they are present across the board in those species. The degree to which these examples demonstrate intentional teaching (i.e., in terms of the teacher's awareness of their facilitatory behaviour) is also questionable. This form of teaching would fit in some functional definitions, such as Caro & Hauser (1992)'s, which states that for teaching to occur, the teacher modifies their behaviour at a cost, only in the presence of the learner. This can include encouragement or punishment resulting from the learner's behaviour, resulting in a newly acquired knowledge or skill. However, just because a behaviour functions to teach, does not mean that the teacher can intentionally deliver it as teaching. Therefore, this would not fit in more cognitive definitions of teaching, such as the '*three steps to recognizing teaching*' as outlined by Davis-Unger & Carlson (2008), who state that to intentionally teach, the teacher has to recognize (a) a gap in the learner's knowledge, (b) that the gap occurs between the teacher and the learner (i.e., you, as the teacher, know more), and (c) that teaching will bring about learning.

The distinction between human and non-human teaching is not in the absence or presence of it, but as suggested by Burdett, Dean, & Ronfard (2017) and Caldwell, Renner, & Atkinson (2018), in specific differences in the execution. Premack (2007) argued that the behaviours seen in the meerkats and the tandem running ants reflect adaptions to specific behaviours, aimed at a single target, that do not generalize to other contexts. In agreement, Thornton & Raihani (2009) stated that human teaching is flexible and generalizable, and that the capacity to do so requires some assumptions to be made about the knowledge of others through an understanding of their mental states. This allows humans to assess ignorance in others, generalize teaching so it becomes useful across contexts, and alter their teaching styles according to the learner's ability. The use of flexible teaching in humans has shown to be useful in transmission studies. For example, Caldwell et al (2018) showed adult humans a selection of knots which they were asked to replicate, then teach what they had learned to a naïve participant. It was found that conditions in which the learner was actively taught were significantly more successful than conditions where the learner was given an end-state or intermediate-state of the cultural parent's performance to independently learn from. While this study is not a test of culture accumulating, it does demonstrate that teaching enables novel skills to be transmitted effectively, particularly skills which the learner might find difficult to uptake. The ability to tailor a teaching style means that new skills are passed on efficiently without the risk or cost to cumulative culture associated with trial and error learning.

Intentional bidirectional communication

Explicit representations of the self and other minds have been proposed to exist, in part, for the purpose of sharing information with others, and to allow this constant updating in line with group members (Shea et al, 2014). Of course, if cultural parents are expected to tailor information to the needs of the cultural offspring, it makes sense to assume the cultural parent would benefit from some cultural offspring feedback. This fits with the most basic model of synchronous teaching, in which the cultural parent and cultural offspring communicate bidirectionally. Feedback from the cultural offspring may assist the cultural parent in fine-tuning the information they provide, such that it becomes more tailored to the needs of the cultural offspring. As such, it is not only the cultural parent's responsibility to select useful information to transmit to the cultural offspring, but the cultural offspring's responsibility to select information about their own knowledge in order to guide the cultural parent's information transmissions.

To effectively communicate about their own knowledge, the cultural offspring must consider what they know relative to what the cultural parent knows, that is, mental state understanding in at least the second order. This means that the cultural offspring has to understand the cultural parent's mental state in order to select information that is informative about their own knowledge, but particularly to the teacher. Cultural offspring communications may serve to clarify to the cultural parent that the cultural offspring knows about a certain thing, but also to indicate that they need more information.

There is a very limited pool of experimental evidence that supports the necessity of bidirectional communication for accumulation of information over generations, however there are a select few studies that do suggest that interactive teaching can support more accurate transmission of information. For example, Caldwell et al. (2018) found that interaction with a teacher did facilitate subsequent learner success in cases where the goal was particularly difficult to achieve, suggesting that interactive teaching may be particularly supportive for hard-to-learn information. Further support comes from experimental literature suggesting that being a learner as part of a conversational dyad (engaging in bidirectional communication) holds significant benefit compared with a learner who is not part of the conversation, but holds the same goal (an '*overhearer*') (Clark & Wilkes-Gibbs, 1986; Schober & Clark, 1989). Because overhearers are not part of the '*grounding*' process that the conversational dyad build, mutual knowledge, or mutually understood signals that act as partner-specific conversational aids do not hold the same benefit (Brennan & Clark, 1996; Garrod & Anderson, 1987). Therefore, there is a clear advantage to cultural offspring who can give feedback to cultural parents about their own performance, or communication errors, or gaps in understanding, relative to overhearers whose specific needs have not been accommodated in the same way.

This section summarised three capacities that are likely to be enhanced or enabled by mental state understanding. These capacities (flexible use of social learning strategies, intentional teaching and intentional bidirectional communication) are all possible benefits for cumulative culture, such as allowing careful selection of social models, and careful selection of tailored information transmissions. This thesis aims to explore the relevance of these capacities further, therefore the following section will assess current tests of cumulative culture that can be modified to achieve this.

1.8 Investigating the effects of uniquely human social cognition on cumulative culture

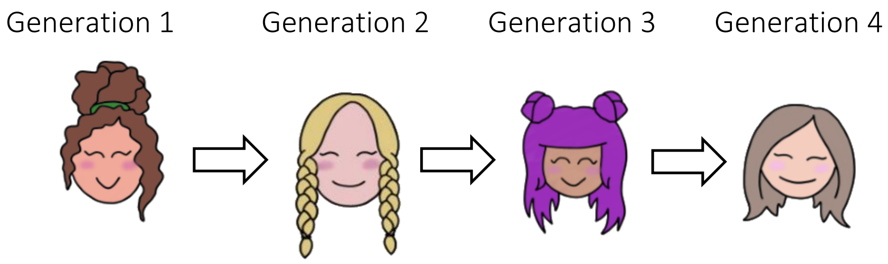
The broad aim of this thesis is to explore the facilitatory effect mental state understanding has in cumulative culture. This is done by altering aspects of traditional lab-based tests of cumulative culture to assess conditions in which information sending is restricted and therefore participants may have to apply some reasoning about other's mental states to choose what to transmit. The following section will discuss the existing methods for testing cumulative culture under lab conditions and provide suggestions as to how these can be adapted to answer the questions in this thesis.

We can observe the effects of cumulative culture in the real world in behaviours and artefacts that take many generations to refine, such as languages and traditions. However cumulative culture is also observable within single biological generations, for example, recent advances in technology in which cumulative culture moves relatively fast. Despite the visibility of cumulative culture in the real world, lab based experimental tests are required to assess the conditions under which cumulative culture can occur in humans, and to explore the supporting capacities that are unavailable to children and non-humans.

Assessment of cultural change is made possible in generational designs, which aim to mimic real-world replacement using experimental learner generations of participants in a lab. These designs can be linear, whereby information is transmitted between individual participants, each representing a single generation, or group focused where the successive replacement of individuals in the group captures cultural continuity, as well as (gradual) generational replacement. In linear transmission chains (Figure 4), information is typically transmitted unidirectionally between individual participants, each of whom represents a separate generation. This method has been widely used to study the transmission of linguistic material, either textual or spoken. Designs of this type have identified preferential sending of certain types of information, such as negative story events over positive story events (Bebbington et al., 2017) and social over non-social information (Mesoudi et al., 2006). Linear transmission chains also demonstrate how over generations, traits tend to evolve such that they become easier for the following generation to learn (Ravignani et al., 2017; Tamariz & Kirby, 2015). Replacement designs also involve transfer from earlier to later generations. However, in replacement chains, the longest serving member of the group is removed and replaced with a new group member at regular intervals in order to mimic the replacement of cultural generations in the real world (Figure 5). This experimental structure has led to replacement designs sometimes being referred to as '*microsocieties*' due to their likeness to real-world societal structure (e.g., Caldwell & Millen (2008a)). Comparisons of conditions under which cumulative culture occurs largely comes from replacement design studies. This is primarily because the continuity generated by the overlapping learner generations protects against loss of information, and the opportunity to learn from more than one individual allows for success-biased copying, both of which are likely to be conducive to the accumulation of beneficial traits. These studies have been implemented in comparisons of conditions of emulation, imitation and intentional teaching (Caldwell & Millen, 2009; Morgan et al., 2015; Zwirner & Thornton, 2015).

Figure 4

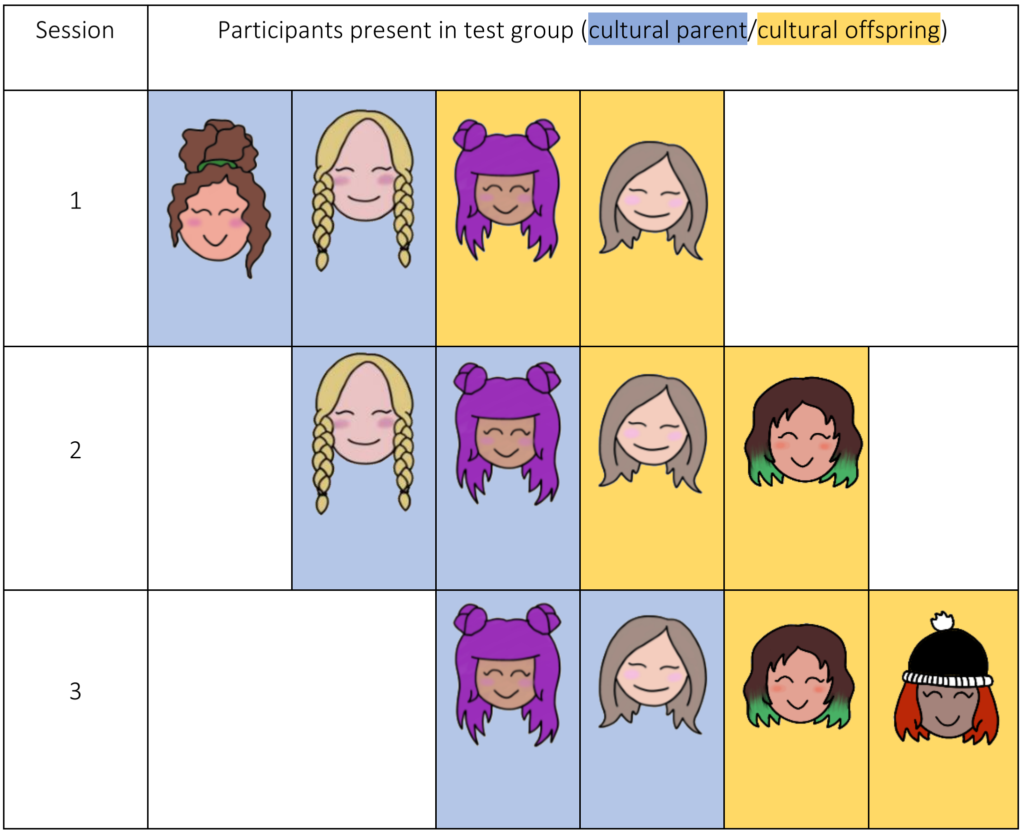
An example of a transmission chain layout



Note. At each generation, a task is completed then some information about the outcome is unidirectionally transmitted to the next generation. Direction of transmission is indicated by the arrows.

Figure 5

An example of the replacement chain layout



Note. A task is completed by the most experienced member(s) shown here with a blue background. Learners are represented by characters with yellow backgrounds. As the sessions progress, more experienced members leave, and learners take their place. New learners are introduced in each session.

Given the success of generational experimental designs in exploring cumulative culture under different social learning conditions, it makes sense to use these as the basis for answering questions about the role of mental state understanding. However, in contrast to most previous studies of the mechanisms involved in cumulative culture, the studies in this thesis will use a more abstract computer-based task which offers greater potential for precise manipulation and measurement.

1.9. Thesis goals

In this review, I have summarised the current literature linking mental state understanding as a facilitator of cumulative culture, identifying an absence of empirical studies directly testing this. I will now outline how the studies in this thesis address this gap using a blend of traditional methods for assessing cumulative culture in the lab, and abstract computer-based tasks.

Chapter 2 of this thesis explored a new methodology proposed by Caldwell et al. (2020), which uses simulated transmission chains to gain a proof of principle that participants can utilise information from a social source in order to increase the rewards available in later generations. This task does not require participants to select any information for transmission, and so we opted against using traditional transmission chain methods. Following on from this, Chapter 3 adopted a computer based linear transmission chain design to assess what information adults transmit to later generations in conditions where information transmission is severely limited. This intentional selection of information was directly compared to a condition in which a subset of inadvertent information was transmitted, and a third condition in which all of the information available was transmitted. Chapter 4 applied the same methodology used in Chapter 3 but instead uses a large developmental sample of 5- to-10-year-old children with the aim to explore the point in development that children begin to perform as adults do in this task. The intricacies of adult intentional information sending are explored further in Chapter 5. This final experimental chapter explored how feedback from an information receiver affects the quality of information transmitted by the producer. Together these studies contribute to the growing literature on cumulative culture in humans and the mechanisms that support aspects of cumulative culture that are not available to other species.

Chapter 2: Demonstrating a new methodology for examining performance improvement over generations of social transmission…without social transmission

2.1 Abstract

Experimentally examining cumulative culture and the mechanisms that support it requires some analysis of the effect of transmission over multiple generations. Previous laboratory research in cumulative culture has typically done this by studying the effect of transmission between multiple participants. Such experimental studies have been used to mimic cumulative culture and thereby investigate some of its supporting capacities in populations of human adults. However, creating large-scale comparisons of different capacities, or comparing the performance of different population groups, can push these designs to their limits – requiring large participant pools and complex organisation that is impractical and prone to error. The current study pilots a new experimental methodology proposed by Caldwell et al. (2020) that aims to address these limitations, while still producing multi-generational data. We did this by assessing twenty adults' ability to adopt a low-risk strategy of repeating behaviour vicariously revealed to be rewarded, and favouring individual exploration over repeating behaviour vicariously revealed to be unrewarded. By presenting information representing every potential reward outcome, in a task involving a graded reward structure, we were able to create a large dataset detailing how each participant would perform in response to exposure to social information from high-, low-, or middle- performing demonstrations. This could then be used to simulate how performance would change over generations of social transmission. The results of this study suggested that this methodology could offer an effective alternative for assessing the potential for cumulative cultural evolution in human and non-human populations. Furthermore, we were able to perform an experimental manipulation that was expected to impact on the rate and extent of cultural evolution, and the methodology used successfully captured this. This

shows that the method is suitable for detecting group-level effects on cultural accumulation under different conditions, thus offering potential for investigating factors implicated in cumulative culture.

2.2 Introduction

Cumulative cultural evolution is widespread in human adults. Claims of its distinctiveness abound because, although it occurs in all known human populations, there is little to no evidence of it in natural populations of non-humans. This has further led to great interest in understanding the mechanisms underlying cumulative culture, and this interest has fuelled experimentation on the topic. Because cumulative culture relies on change over generations, investigating it necessarily requires methods that capture this generational change. This has typically involved using multi-generational transmission chains, which can be challenging in some populations due to practical limitations, such as participant organisation. In the current study, we test a new method that is capable of producing datasets that allow investigation of cumulative culture without depending on the transmission of information between individuals.

The current study aims to test the '*potential for ratcheting*' ('PFR') principle outlined by Caldwell et al. (2020). We discuss the issues with current methods for measuring cumulative culture, and introduce this innovative alternative methodology which addresses many of the limitations of the existing methods. To ensure that this novel method is viable however, it is important to establish proof of principle, prior to attempting to use it to test hypotheses about the presence or otherwise of cumulative culture under different conditions. Our decision to pilot this methodology using adult participants was motivated by the uncontroversial expectation that human adults possess the capacities that this method is designed to assess, meaning that we can reasonably expect to produce the signatures of cumulative culture in this population. These results will provide an important baseline for comparison with other populations, facilitating interpretation of any negative, or more equivocal results found in other groups for which capacities for cumulative culture are currently debated.

Current experimental tests of cumulative culture

As aforementioned in Chapter 1, cumulative cultural evolution is the capacity that allows for improvements to behaviours and artefacts as a result of repeated transmission over generations of learners. Its characteristic ratchet effect (Tennie et al., 2009) means that transmission results in changes that are typically in the direction of improvement, with little to no reversal. Current tests of cumulative culture aiming to capture this effect rely on multi-generational transmission studies which mimic real-world replacement using sample experimental generations of participants (Caldwell, 2018).

A variety of methods for generating this multi-generational transmission have contributed to not only demonstrating cumulative culture in adult populations, but also going some way to unravelling the cognitive mechanisms required for it to occur:

'Linear transmission chains' involve the unidirectional transfer of information from one individual (referred to throughout this thesis as the '*cultural parent'*) to another (the '*cultural offspring'*). There are few examples of this method being used to capture cumulative culture, however there are several studies demonstrating related phenomena, suggesting that it could, in principle, capture cumulative culture. For example, this method has been used to demonstrate preferential sending of certain types of information (Bebbington et al., 2017; Mesoudi et al., 2006) and the changes to information as a result of repeated transmission (Ravignani et al., 2017; Tamariz & Kirby, 2015).

Similarly to linear transmission chains, '*replacement designs'* (sometimes termed '*microsocieties*') also involve transfer from earlier to later generations. However, membership of adjacent generations overlap. The longest-serving members are removed and replaced with new group members at regular intervals in order to mimic the process of generational replacement in the real world (e.g., Caldwell & Millen 2008a). This method has also been implemented in lab conditions with adult participants to test social learning mechanisms necessary for cumulative improvement to occur (Caldwell & Millen, 2009; Zwirner & Thornton, 2015). Related capacities have also been explored using this methodology, such as the emergence of traditions (Baum et al., 2004).

General issues in using multi-generational transmission chain designs

While tests of multi-generational transmission are proficient for establishing the outcome of repeated transmission, and allowing scope for manipulation to isolate the cognitive mechanisms involved, there are several practical limitations that make them a less than ideal task design. Recreating multi-generational transmission requires participants to be coordinated to either be in the same place at the same time, or at staggered time intervals. Such coordination is difficult to control, so for example, in a testing circumstance whereby participants have to meet at a particular time and location to take part. This issue becomes even more significant when considering the large participant pool required in order to run a large number of chains, which will be required in studies with several conditions. Examining multiple transmission conditions, along with appropriate experimental controls, creates a significant increase in the number of participants required. Some examples include Zwirner & Thornton (2015) who used a total of 190 participants for ten chains of six participants in each of three conditions, Tamariz & Kirby (2015), who used a total of 308 participants for eight chains of 22 in each of two conditions, and Caldwell & Millen (2009) who used a total of 700 participants for ten chains of ten participants in each of seven conditions. It is perhaps also worth noting that despite these large participant pools, some of these studies still failed to find significant differences between conditions where effects had been predicted (Caldwell & Millen, 2009; Zwirner & Thornton, 2015), and it remains possible that the recruitment numbers, although high, were still insufficient to detect effects that might have been captured with greater statistical power.

The issues of participant numbers and participant organisation become even more prevalent when assessing groups of children or non-humans, as discussed in depth in Caldwell et al. (2020).

Simulating multi-generational transmission

The current study attempts to provide a proof of principle for a flexible methodology outlined by Caldwell et al. (2020), which reduces the need for transmission chain designs and the issues associated with their implementation. This PFR (‘*potential for ratcheting’*) methodology states that transmission chains can be simulated by collecting and arranging participant’s responses from multiple transmission outputs of varying quality, mimicking performance at different generations (as opposed to traditional chain tasks which generally only collect one response, following exposure to one transmission output). We used a grid search task in which cultural offspring (henceforth in this chapter, ‘*participants’*, since all played the role of offspring in the current task) were given vicarious information from a computer-generated parent model about what could be found hidden under some of the grid squares.

Following the PFR methodology, participants were shown multiple sets of vicarious information on a parent model. Participants were required to search the grids to find the rewards using the vicarious information they were given to help them. As all grids included vicarious information about some of the searchable locations, participants always chose to repeat or avoid selections made by the parent model. The full set of trials given to each participant covered every possible parent grid score that could be achieved by making the required number of selections in that grid. Collecting participants’ scores from these sets of vicarious information allowed assessment of whether participants could reliably outperform a parent grid at a range of different levels of difficulty. From these transitions (parent model score = X; participant score = Y) it is then possible to infer the outcome of a linear transmission chain of a series of participants.

The score data from participants was used to simulate artificial chains of participants in which task scores were matched to the vicarious information given to another participant. For example, the first position in the chain would be formed by the behaviour of participant A. If participant A observed a parent grid that scored three points and subsequently scored five points when searching their own grid, the score for position one in the chain would be five points. The score for position two in the chain would then be chosen from the behaviour of another participant who had observed a parent grid that scored five points, and so on until the desired number of chain positions is complete.

To our knowledge, only one study has attempted to simulate transmission chains from individual behavioural data in this way. Claidière et al. (2018) simulated chains using data from an earlier study where baboons were trained to reproduce patterns of grid squares on a screen (Claidière et al., 2014). While this earlier study was not a test for the potential for ratcheting in the baboon population, it did produce outputs that could be used to simulate chain positions. The later study concluded that the simulated data not only showed an increase in reproducibility (i.e., participants created more clusters of information than their predecessors) over generations, but also that it mirrored the real chain data in the earlier study. This suggests that a simulation task assessing the potential for ratcheting could potentially provide fairly accurate predictions about what would happen in inter-individual transmission. Furthermore, it shows that this kind of task may be applicable for use in non-humans and extended into tasks for adults and children for comparative purposes.

The overarching aim of the current study is to demonstrate the efficiency of the PFR paradigm by applying it to a small population of human adults (as aforementioned, a population of which the uncontroversial expectation is held that they possess the capacities that this method is designed to assess). This aim is extended to providing a basis for future research in children and non-humans using the same paradigm.

We explored the PFR principle in two contexts. The first context represented circumstances in which the vicarious information revealed high-precision payoff information. In this condition only, information about the specific locations in which the rewards were found was available, as well as the search pattern and the overall score for the search. Participants’ lowest-risk performance in this condition would be driven by rational decisions about which of the model’s selections to retain, and which to disregard. Therefore, in this condition where participants were aware of the outcome of individual parent selections in situations where vicarious information was of high precision, they could select stimuli revealed to be rewarded (successfully repeating), avoid stimuli revealed to be rewarded (omission error), select stimuli revealed to be unrewarded (commission error) or avoid stimuli revealed to be unrewarded (successfully avoiding). As such, this condition captures the outcome of cumulative culture when the learner has insight into the effectiveness of individual elements of a cultural trait, and can access which elements to retain as a result of causal reasoning (Osiurak et al., 2016). This context was the key in terms of establishing a proof of principle for the potential for ratcheting task, because the link between the behaviour of the cultural parent and the payoffs of that behaviour are clear, and so outperformance of the parent model was expected.

The second context represented a low-precision payoff setting whereby the search pattern and overall score for the grid was visible, but the specific locations of the rewards were not. In this context, the link between parent behaviour and payoffs is much harder for participants to establish. In this condition, scores can be maximised by making a rational decision about whether to repeat the entire model performance, or to disregard everything and rely on individual exploration, based on a threshold score (which varied according to the specific dimensions of the grids). Therefore, in this condition, where participants were not aware of the outcome of individual parent selections where vicarious information was of low precision, they could choose to repeat all parent selections or avoid all parent selections. However, for scores in this condition to accumulate across the chains, this relied upon either (extremely rare) occasions when the '*avoid-all*' strategy resulted in selecting all of the rewarded locations by chance, or the use of a (suboptimal, on average) '*mixed*' strategy, whereby some of the model selections were repeated and others were avoided, based on guesswork. This condition captures situations in which cumulative culture is not insightful, but rather emerges through the preferential spread of adaptive traits as a result of wholesale copying of successful or prestigious individuals, combined with variation contributed by random copying errors and blind exploration (Derex, Bonnefon, Boyd, & Mesoudi, 2019).

Both conditions represent realistic learning scenarios in which an observer must make an executive decision about how to use available information to maximise their own rewards. For example, in a scenario where a hungry forager must climb trees to search for hidden fruit following a previous forager who returned with armfuls of fruit, the High-precision condition represents circumstances in which the previous forager's search is indicated by both fallen leaves (indicating searched locations) and parted branches revealing the trees' bounty. In the same foraging scenario, the Low-precision condition represents circumstances in which the previous forager's search is indicated by fallen leaves, but no information about the specific locations of their rewards is available because no branches have been parted.

In the current task, participants had the goal to select as many rewards on the grid as possible using the vicarious information from a parent model to assist them. In the High-precision condition, individual participants could increase the likelihood of succeeding in this by adopting a strategy whereby rewarded behaviour is repeated, and unrewarded behaviour is avoided in favour of individual exploration. While a low-precision setting might not allow such careful planning, adopting this insightful strategy in a high-precision setting would guarantee matching or outperforming the parent model every time. Therefore, in the High-precision condition, we expected to find a strong preference for repeating rewarded behaviour and avoiding unrewarded behaviour, in line with the goal to maximise the score. In keeping with this prediction, we also expected simulated chains to show the steady increase in scores over generations until the maximum score is reached, after which we expected to see a plateau effect, with no decline.

While we had no certain expectations about how participants would behave in the Low-precision condition, there was a limited number of ways that they could respond to the information: to repeat all of the selections, avoid all of the selections, or to repeat some selections and avoid others in an arbitrary fashion (mixed strategy). The low-risk strategy for individuals would be to adopt a conditional '*repeat all selections*' or '*avoid all selections*' strategy, depending on the number of rewards uncovered in the parent grid. Vicarious information showing an above-chance performance should be repeated fully, and vicarious information showing a below-chance performance should be avoided fully. While the '*repeat-all*' strategy would maintain any uncovered rewards, adoption of the '*avoid-all*' strategy, particularly in cases where the parent model shows no rewarded selections, may present an opportunity for all of the rewards to be uncovered by chance.

There was an expectation that this '*repeat-all*' or '*avoid-all*' strategy would be observed in at least some of the participants, some of the time. It was expected that participants would also adopt the mixed strategy occasionally. While it is not possible for participants using a mixed strategy to outperform the low-risk repeat-all/avoid-all strategy as perfectly applied, it may also increase the opportunities to select all of the rewarded stimuli by chance. We expected simulated chains in this condition to reveal much more limited accumulation of rewards, compared with the high-precision setting.

We also included a risk questionnaire adapted from Weber et al. (2002), to assess whether participant's likelihood of avoiding parent selections and taking their chances of finding the rewards individually correlated with their likelihood to take risks in every-day scenarios. While we had no certain predictions about how this test would predict behaviour in the Low-precision condition, we expected that the most risk-prone participants would also be most likely to avoid parent selections and take their chances in individual exploration.

2.3 Methods

2.3.1 Participants

The final sample consisted of 20 adult participants aged between 17 and 37 years (median = 20, 16 females). Two additional participants were recruited but withdrew from the study before completing the task. All participants in this sample were recruited from the University of Stirling via an online recruitment database through which students can earn tokens required for course completion by participating in research studies.

The study was approved by the University of Stirling General University Ethics Panel (approval reference number: GUEP 35). Every participant gave written consent before the experiment took place.

2.3.2 The experimental task

The task completed by participants was a landscape search task run on desktop computers. The task was run using Javascript online, with jsPsych (de Leeuw, 2015).

Participants were presented with search landscapes (‘*grids*’) containing multiple rows of randomly coloured stimuli (‘*marbles*’) (Figure 6). Each row in the grid contained one target marble and several non-target marbles. Selecting the target marble scored one point. Selecting any of the non-target marbles in the same row, scored no points. Points gained cumulatively over all rows in the current grid were documented using a score tracker positioned above the grid.

Selected marbles were characterised by a blurry version of the previously unselected marble, with a score marker next to it. In the High-Precision condition only, a score marker appeared next to selected targets if the selected marble was a target, the score marker was a green tick, and if the selected marble was not a target, the score marker was a red cross. In the Low-Precision condition, no feedback was provided about the success or otherwise of individual row selections.

Until selected, target marbles were not distinguishable from non-targets by colour or position. The position of the target marble was randomly assigned in each row. Participants were all given the same goal, to select target marbles to score as many points as they could across all grids.

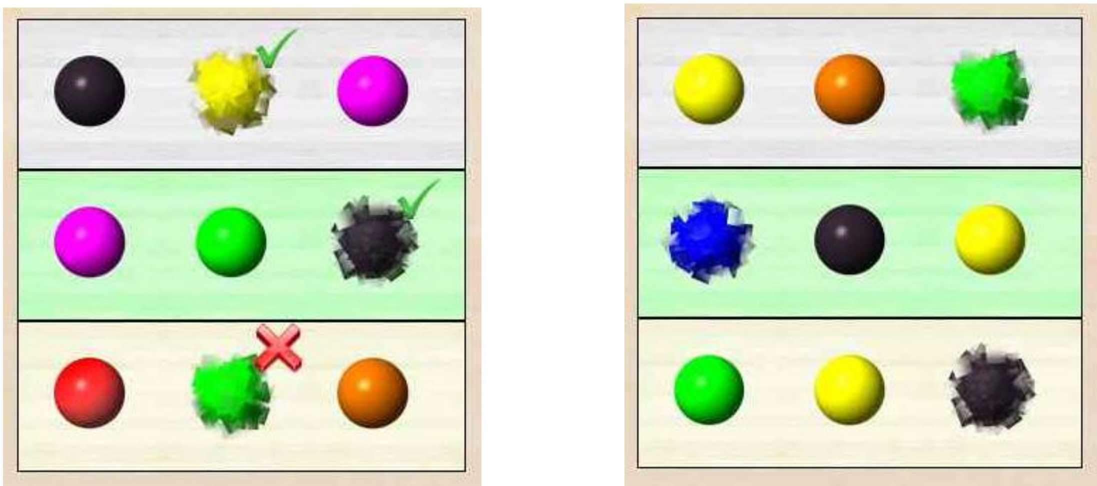
Participants were presented with different sizes of search grid. The smallest contained three rows, each with a width of three. The largest contained six rows, each with a width of six. All possible size variants in between these sizes were used, and therefore 16 different permutations were presented (as shown in Table 2). Grid sizes were shown in a randomised order.

Participants were given some vicarious information on an identical grid shown above their own grid prior to making any selections, which they could use however they wished. This parent grid was visible throughout the trial and was positioned above the landscape on which the participant made their own selections. Despite being computer generated, this information was intended to reflect exposure to a previous participant’s attempt for the same grid, such that the marbles were shown to have been selected. Annotated example screenshots for each condition can be seen in Figure 7a and 7b.

Each participant was presented with all grid sizes showing all possible scores, for example, four 3x3 grids were shown, one of each showing a score of 0, 1, 2 and 3 out of 3 possible points. Therefore, participants were given a total of 88 landscapes to search in each condition (Table 2).

Figure 6

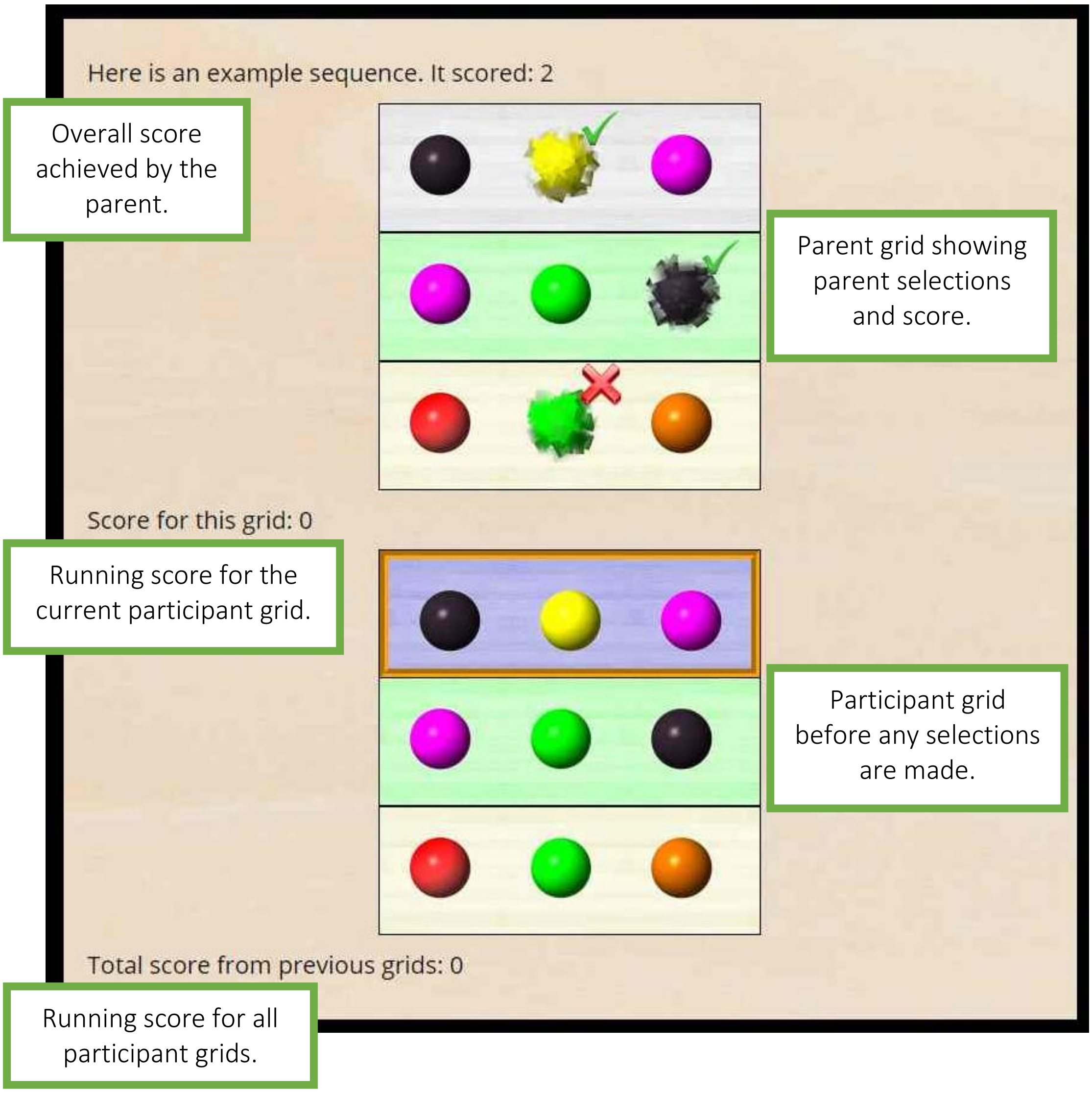
Screenshots of the vicarious information given to participants in the High-precision condition (left) and the Low-precision condition (right).



Note. Participant grids were identical, only without any selections made. Selected marbles are blurred. In the High-precision parent model grids, green ticks indicate that the parent model selected a target marble, and red crosses indicate that the parent model selected a non-target model.

Figure 7(a)

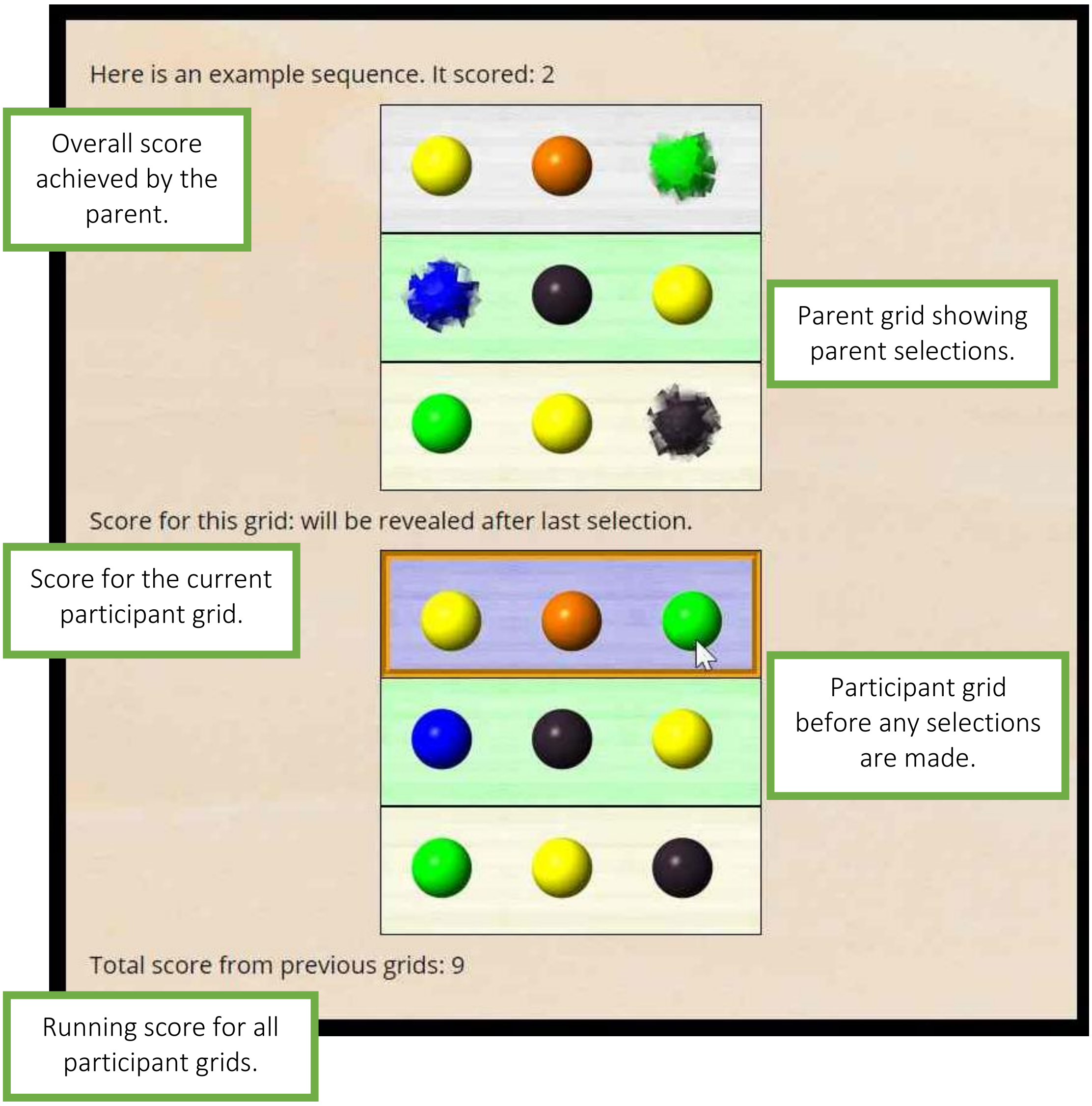
Annotated trial screen from the High-precision condition.



Note. Annotations in green boxes were not visible to participants and are for the benefit of this diagram.

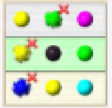
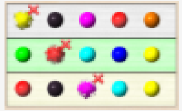
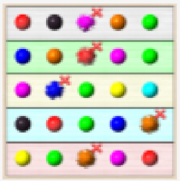
Figure 7(b)

Annotated trial screen from the Low-precision condition



Note. Annotations in green boxes were not visible to participants and are for the benefit of this diagram.

Table 2 *All grid sizes and parent reward variations shown to participants. All grid sizes with one of each variation was shown in each condition.*

Grid Size	Reward Variations Shown	Grid Size	Reward Variations Shown	Grid Size	Reward Variations Shown	Grid Size	Reward Variations Shown
3_3 	0/3 Rewarded 1/3 Rewarded 2/3 Rewarded 3/3 Rewarded	3_4 	0/3 Rewarded 1/3 Rewarded 2/3 Rewarded 3/3 Rewarded	3_5 	0/3 Rewarded 1/3 Rewarded 2/3 Rewarded 3/3 Rewarded	3_6 	0/3 Rewarded 1/3 Rewarded 2/3 Rewarded 3/3 Rewarded
4_3 	0/4 Rewarded 1/4 Rewarded 2/4 Rewarded 3/4 Rewarded 4/4 Rewarded	4_4 	0/4 Rewarded 1/4 Rewarded 2/4 Rewarded 3/4 Rewarded 4/4 Rewarded	4_5 	0/4 Rewarded 1/4 Rewarded 2/4 Rewarded 3/4 Rewarded 4/4 Rewarded	4_6 	0/4 Rewarded 1/4 Rewarded 2/4 Rewarded 3/4 Rewarded 4/4 Rewarded
5_3 	0/5 Rewarded 1/5 Rewarded 2/5 Rewarded 3/5 Rewarded 4/5 Rewarded 5/5 Rewarded	5_4 	0/5 Rewarded 1/5 Rewarded 2/5 Rewarded 3/5 Rewarded 4/5 Rewarded 5/5 Rewarded	5_5 	0/5 Rewarded 1/5 Rewarded 2/5 Rewarded 3/5 Rewarded 4/5 Rewarded 5/5 Rewarded	5_6 	0/5 Rewarded 1/5 Rewarded 2/5 Rewarded 3/5 Rewarded 4/5 Rewarded 5/5 Rewarded
6_3 	0/6 Rewarded 1/6 Rewarded 2/6 Rewarded 3/6 Rewarded 4/6 Rewarded 5/6 Rewarded 6/6 Rewarded	6_4 	0/6 Rewarded 1/6 Rewarded 2/6 Rewarded 3/6 Rewarded 4/6 Rewarded 5/6 Rewarded 6/6 Rewarded	6_5 	0/6 Rewarded 1/6 Rewarded 2/6 Rewarded 3/6 Rewarded 4/6 Rewarded 5/6 Rewarded 6/6 Rewarded	6_6 	0/6 Rewarded 1/6 Rewarded 2/6 Rewarded 3/6 Rewarded 4/6 Rewarded 5/6 Rewarded 6/6 Rewarded

The experiment followed a within-participants design, and all participants took part in both conditions in the same order. The reason that all participants took part in the High-Precision condition followed by the Low-Precision condition rather than a counterbalance, was to ensure they understood how the task worked before moving on to a condition that was necessarily more opaque. The vicarious information for each trial varied depending on the condition:

Condition 1 (High-precision payoff information): Cultural parent grids showed both the position of the selected marbles, and the score marker (green tick or red cross) for each row. As participants selected the marbles on their own grid, they were provided with feedback about the outcome of their selection (target or non-target). Feedback was provided by means of the same score marker that was used in the parent grid. The purpose of this condition was to explore the capability of adults to outperform a model when everything about the outcome of the parent search is known, thus testing potential for ratcheting under conditions intended to make this as straightforward and cognitively undemanding as possible.

Condition 2 (Low-precision payoff Information): Cultural parent grids showed the position of the selected marbles, but no individual row score marker. Participants were given the overall score that the parent achieved for the grid and were only given their own score on completion of all selections in that grid. The purpose of this condition was to explore what strategies participants adopt when it is not possible to selectively repeat only beneficial elements of a behaviour, and to avoid all known failures. As such, this condition tests the PFR paradigm and simulation method under circumstances where the potential for ratcheting is necessarily constrained.

2.3.3 Procedure

Participants were provided with thorough instructions on screen before beginning the experiment. These explained that only one marble from each row would be worth one point, while the others would be worth no points. Participants were instructed to search for target marbles by selecting one marble from each row in each grid. After participants had made all of the selections for a grid, a new grid was shown.

All participants took part in both conditions, completing 88 grids in each (176 trials in total). After completing the first condition, participants could take a short break before continuing to the second condition. An example of the task layout can be seen in Appendix 1.

2.3.4 Additional testing measures

Participants were given a risk behaviour scale consisting of 40 risky scenarios (Appendix 2) in which participants must rate their likelihood of engaging in each. Ratings were made between 1 (very unlikely) and 5 (very likely), with a rating of 0 indicating that the participant would prefer not to answer. The scale used was adapted from an established measure outlined in Weber et al. (2002).

The results of this scale were used to assess how participants' likelihood of avoiding parent selections in the low-precision grids correlated with their likelihood to take risks in real life.

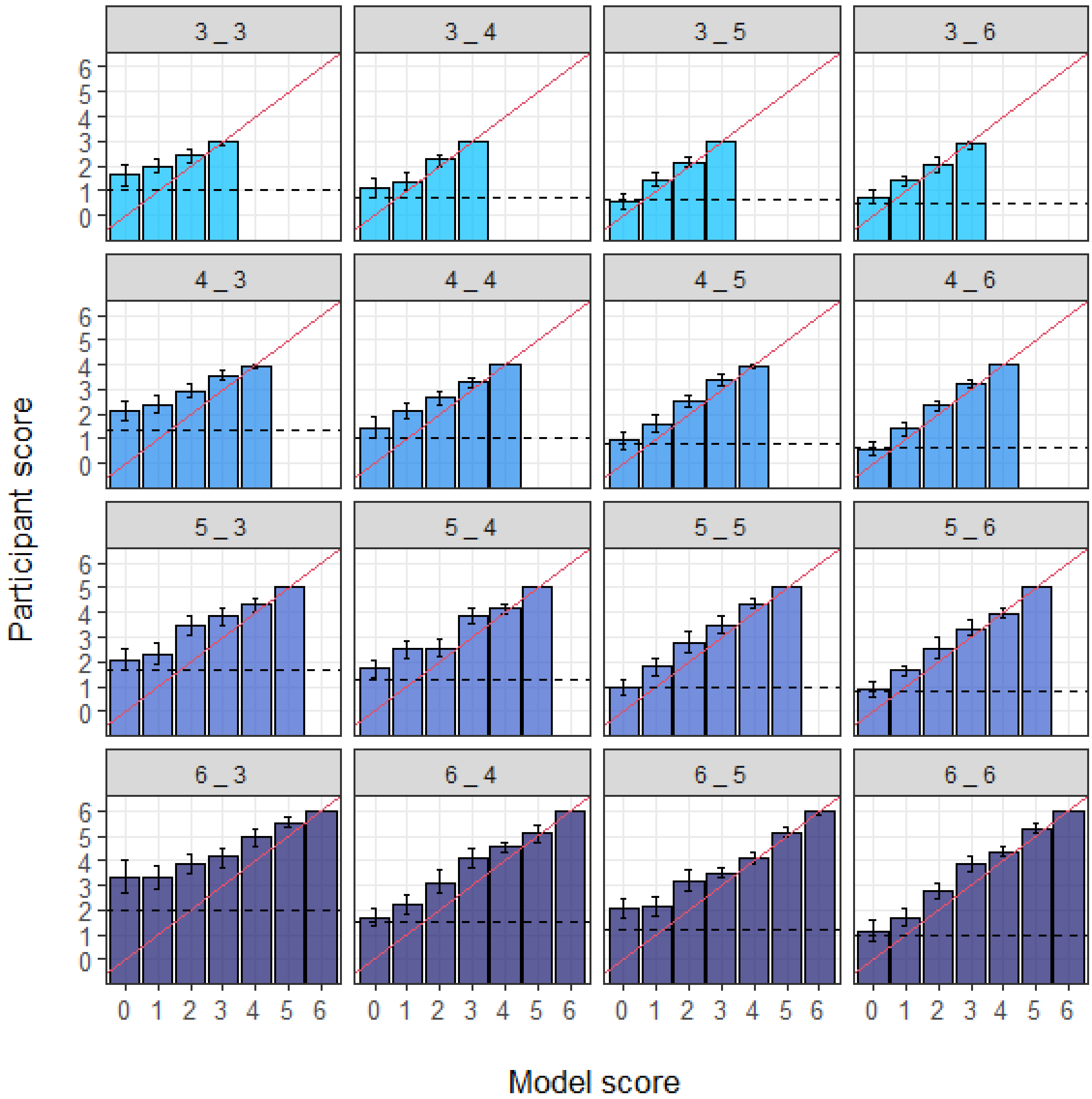
2.4 Results

Graphs were created using ggplot2 performed on R Studio (RStudio Team, 2020), supported by R version 4.0.2.

The mean scores for each landscape size and reward structure in the High-precision condition are shown in Figure 8, and the mean scores for each grid size and reward structure in the Low-precision condition are shown in Figure 9.

Figure 8

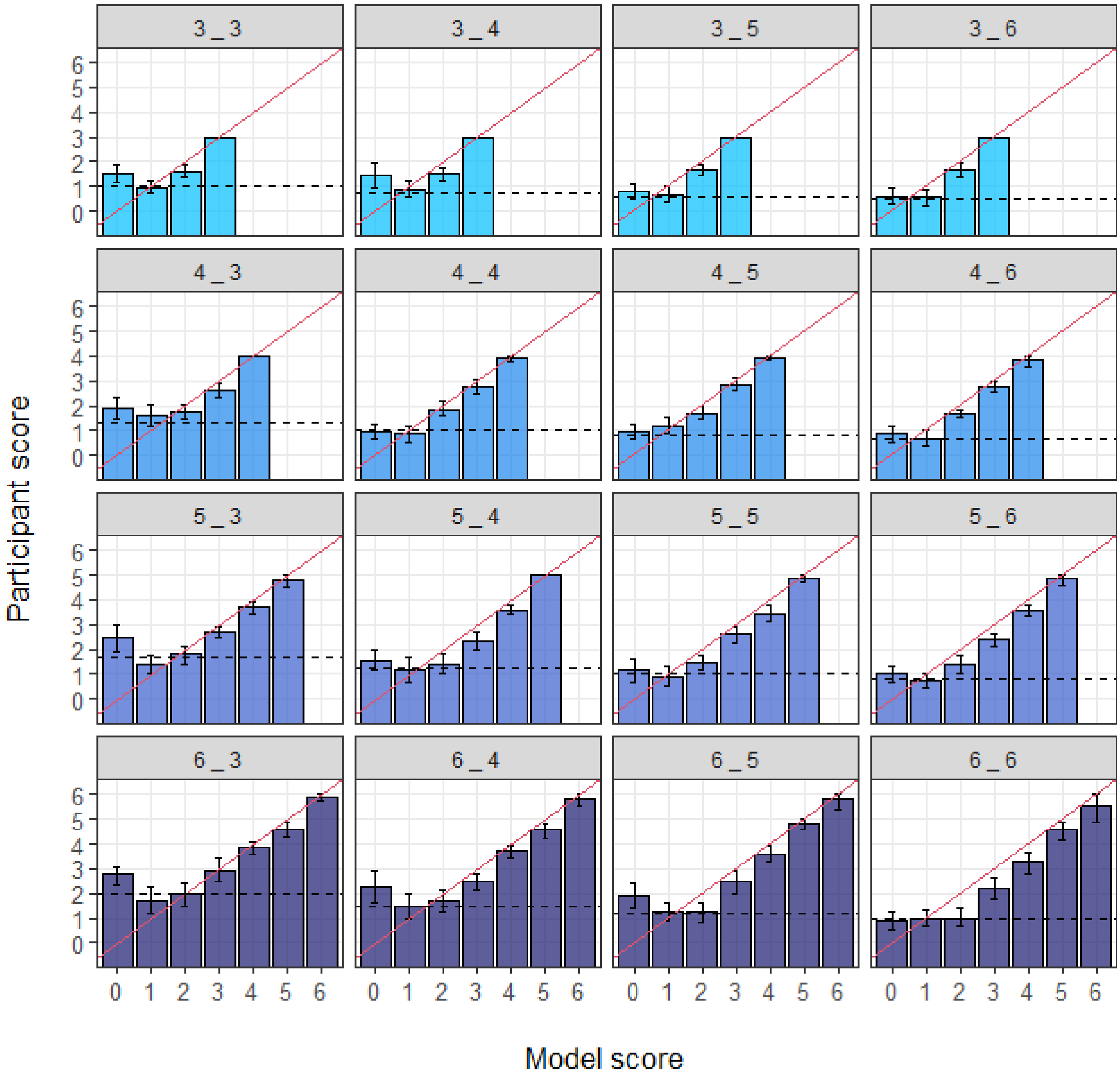
Average participant score in the High-precision condition for each parent score shown, split by grid size.



Note. Grid sizes are labeled as "number of rows" _ "number of stimuli on each row". Error areas represent +/- 1 Standard Error. Red diagonal lines show where the x-axis is equal to the y-axis, i.e., the score achieved by the model, and black dashed lines show the score that would be expected by chance (calculates as (1/number of stimuli per row)*number of rows).

Figure 9

Average participant score in the Low-precision condition for each parent score shown, split by grid size.



Note. Grid sizes are labeled as "number of rows" _ "number of stimuli on each row". Error areas represent +/- 1 Standard Error. Red diagonal lines show where the x-axis is equal to the y-axis, i.e., the score achieved by the model, and black dashed lines show the score that would be expected by chance (calculates as (1/number of stimuli per row)*number of rows).

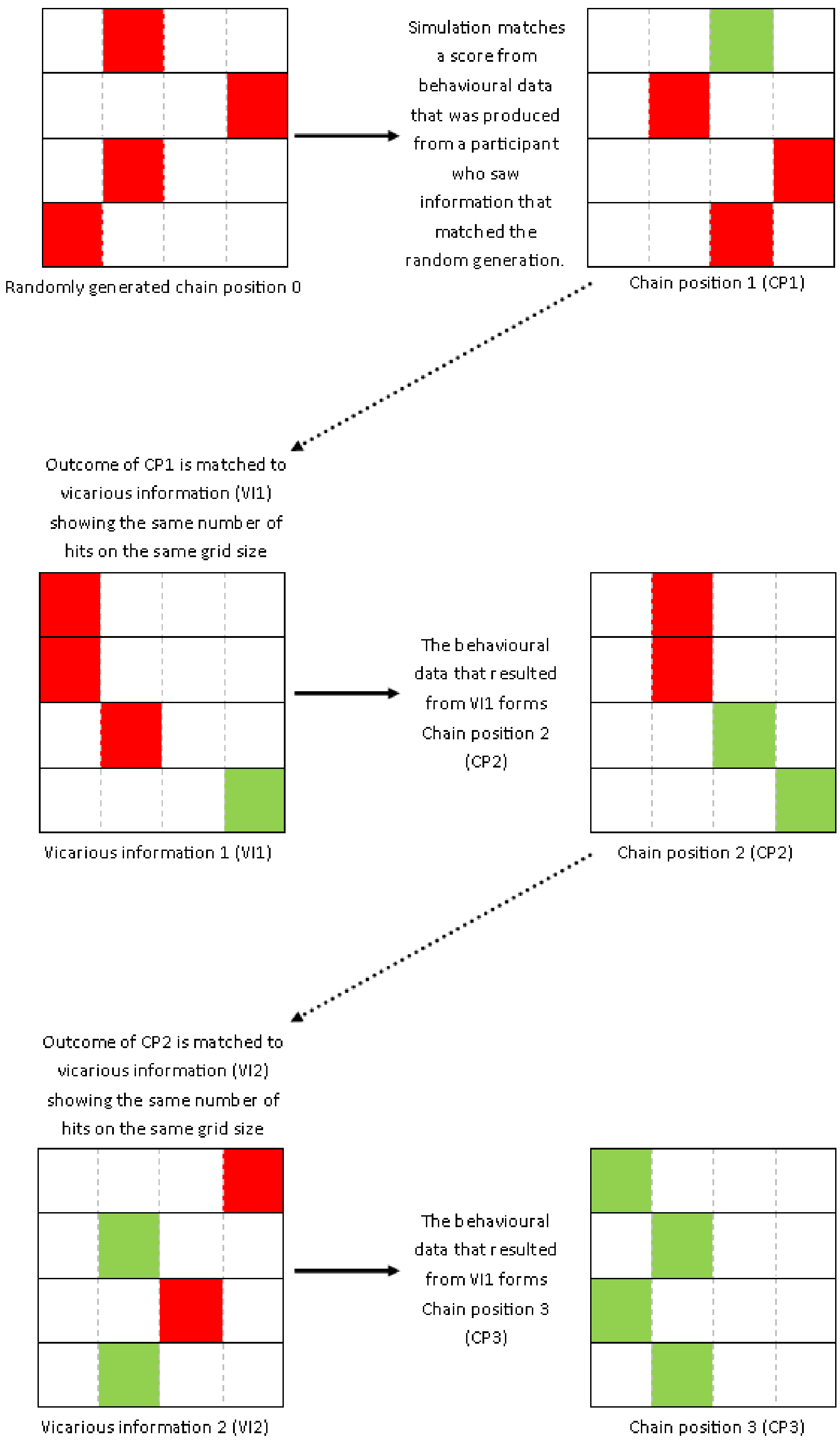
2.4.1 Simulating behavioural data to infer the outcome of a transmission chain

We simulated the data collected in the High-precision condition to demonstrate how participants would have performed if they were part of a generational transmission chain. Simulations were performed using R Studio (RStudio Team, 2020), supported by R version 4.0.2.

At generation 1 of the simulation an agent would receive a score from generation 0 (randomly generated). The score a participant produced in response to a grid of the same dimensions, with the same score, would then form the input for an agent in generation 2. For example, if generation 1 scored three points, generation 2 would be formed using the score of a randomly selected participant that saw vicarious information scoring three points, and so on (Figure 10).

Five hundred simulations were run for each grid type, and the average score for each chain position was taken. The cut-off point for number of generations was decided based on the point at which scores appeared to have plateaued for all grid types. Simulated generation graphs for each landscape size in the High-Precision condition are shown in Figure 11.

Figure 10



Demonstration of how the simulation created the chain data by matching behavioural outcomes to vicarious information.

Figure 11

Average score increase over simulated learner generations in the High-precision condition.

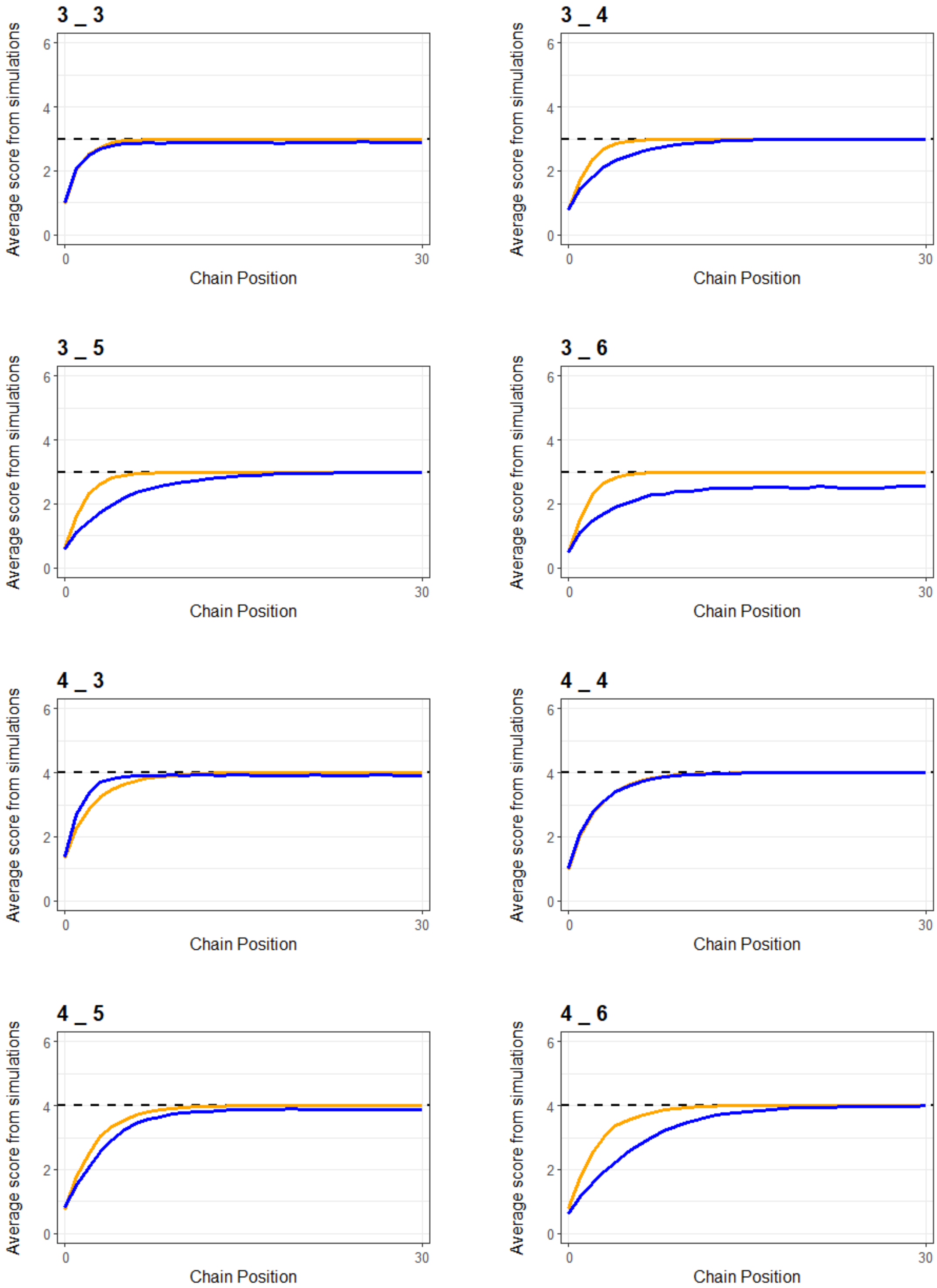
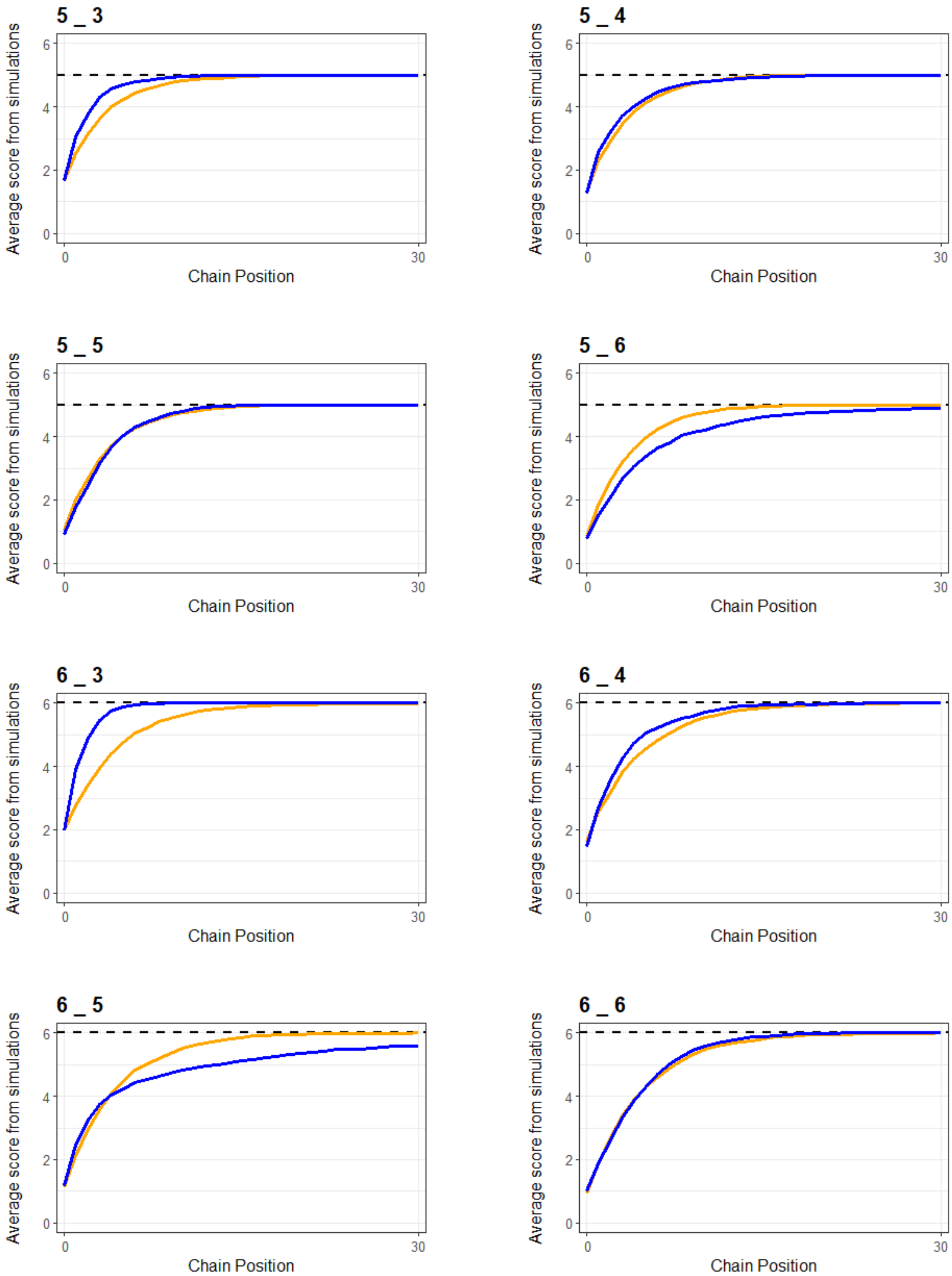


Figure 11 cont.



Note. The average score was taken from 500 separate simulations each run for 30 generations. Grid sizes are labeled as “number of rows” _ “number of stimuli on each row”. Red horizontal lines mark the maximum score for the grid size shown, the dark blue curve shows the simulated data, and the orange curve shows the ‘low-risk *performance strategy’* for that grid, which was calculated by simulating all chain positions to copy all rewarded information, and avoid all unrewarded information.

2.4.2 Assessing the use of the low-risk strategy for maximising low-precision information

In the Low-precision condition, the simulated generations of smaller landscape sizes (up to and including 4x3 landscapes) are similar to the increases seen across the High-precision condition. While they do take several hundred fewer generations to plateau, the increases in the larger high-precision landscapes in the High-precision condition are starkly different from larger landscape sizes in the Low-precision condition, which never peak and instead level off well below the maximum score attainable. Simulated generation graphs in the Low-precision condition can be seen in Figure 12.

In the Low-Precision condition, individuals could adopt a low-risk strategy to maximise their individual score. Doing so involved either repeating all of the selections made when the parent model showed an above-chance score, or avoiding all of the selections made when the parent model showed a below-chance score. Of course, this strategy was optimal for increasing individual performance as opposed to the population goal to increase the number of rewards found over generations. The low-risk strategy that could be adopted for each landscape size (i.e., '*repeat-all'* or *avoid-all'*) was not disclosed to participants.

Figure 13 shows the percentage of participants that opted for each strategy for each size of landscape and parent score. As expected, when the correct strategy was not as obvious (i.e., around the threshold point where the low-risk response shifts between repeat-all and avoid-all), participants generally chose to use a mixed strategy (i.e., repeat some of the model selections but not all of them).

Figure 12

Average score increase over simulated learner generations in the Low-precision condition.

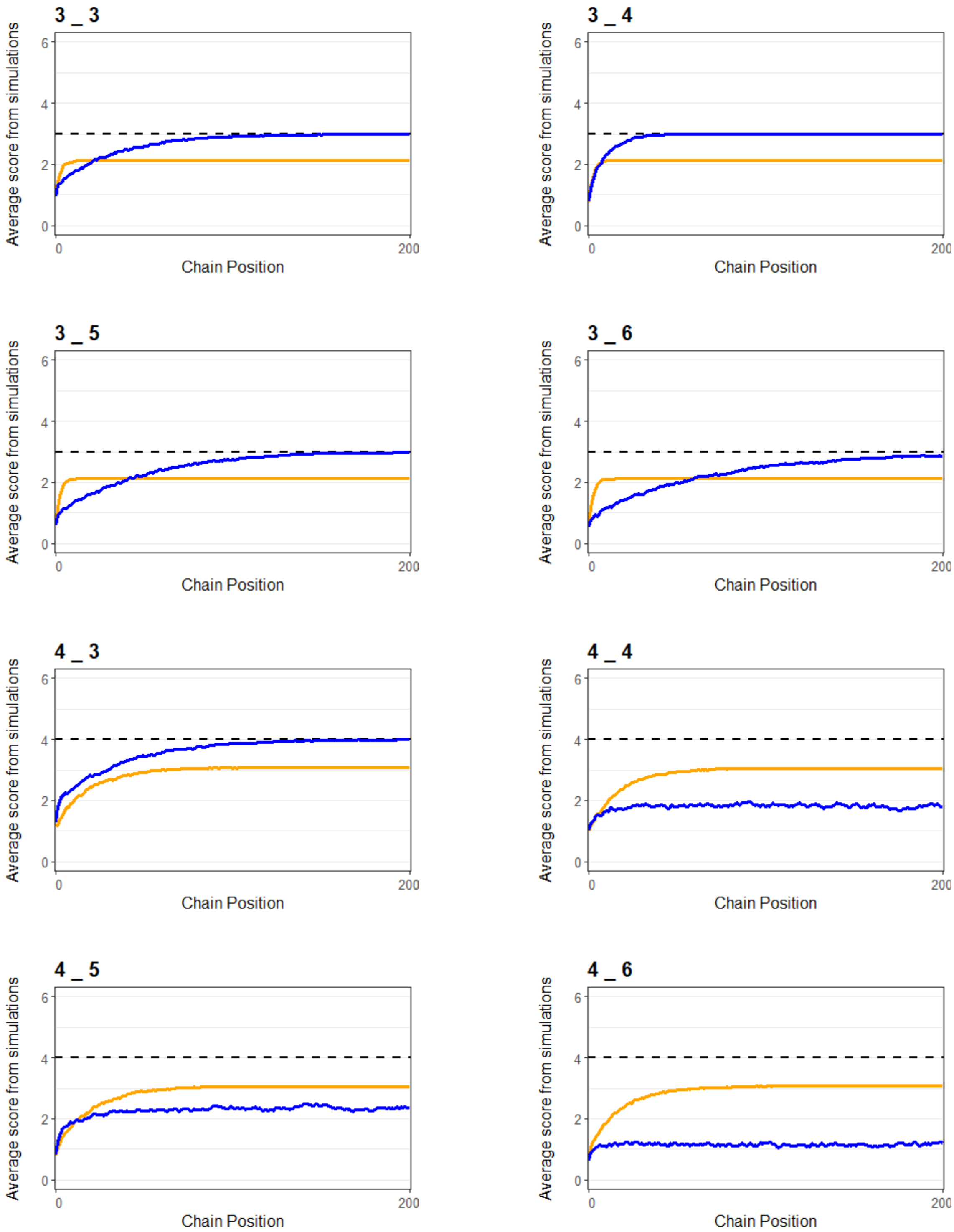
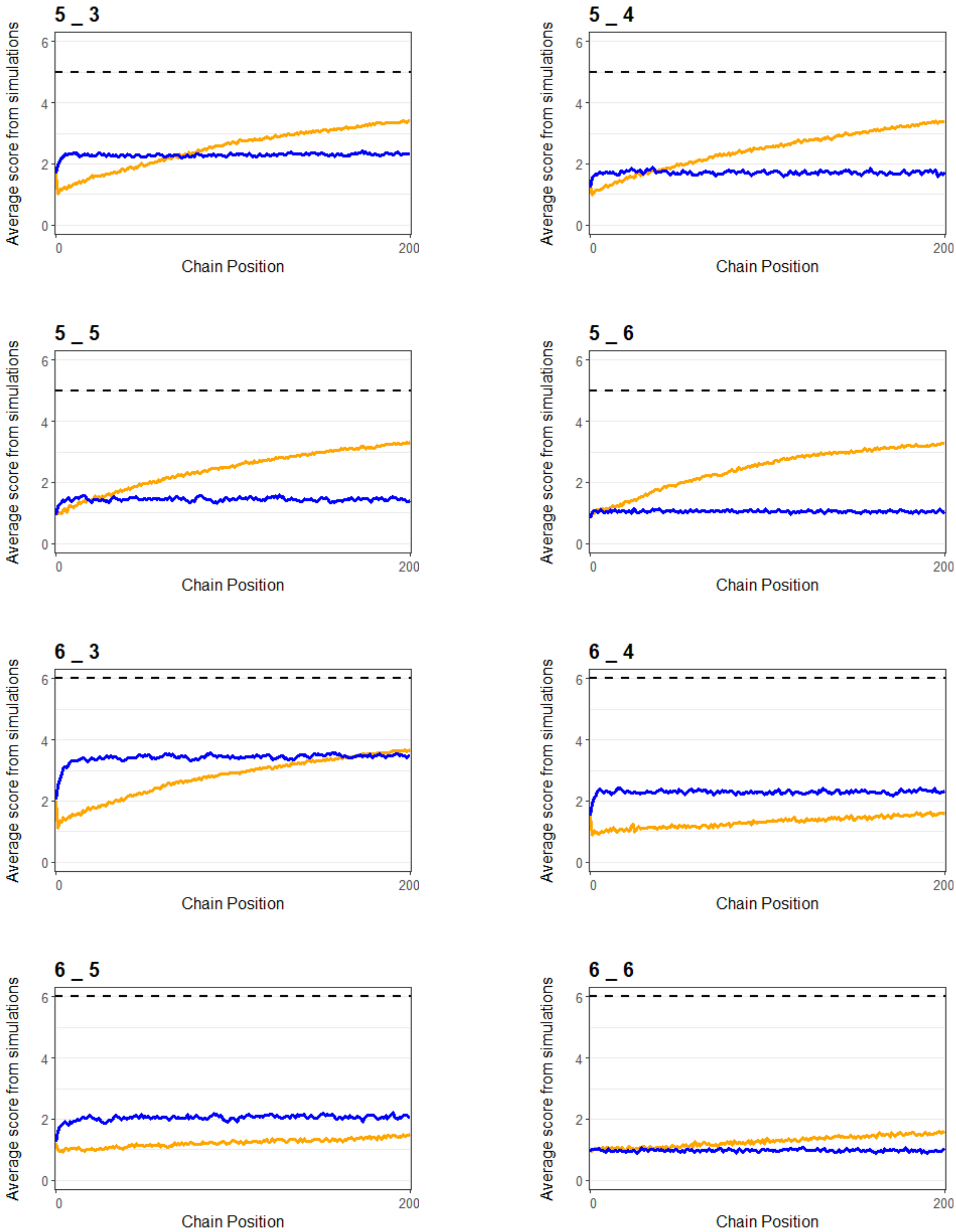


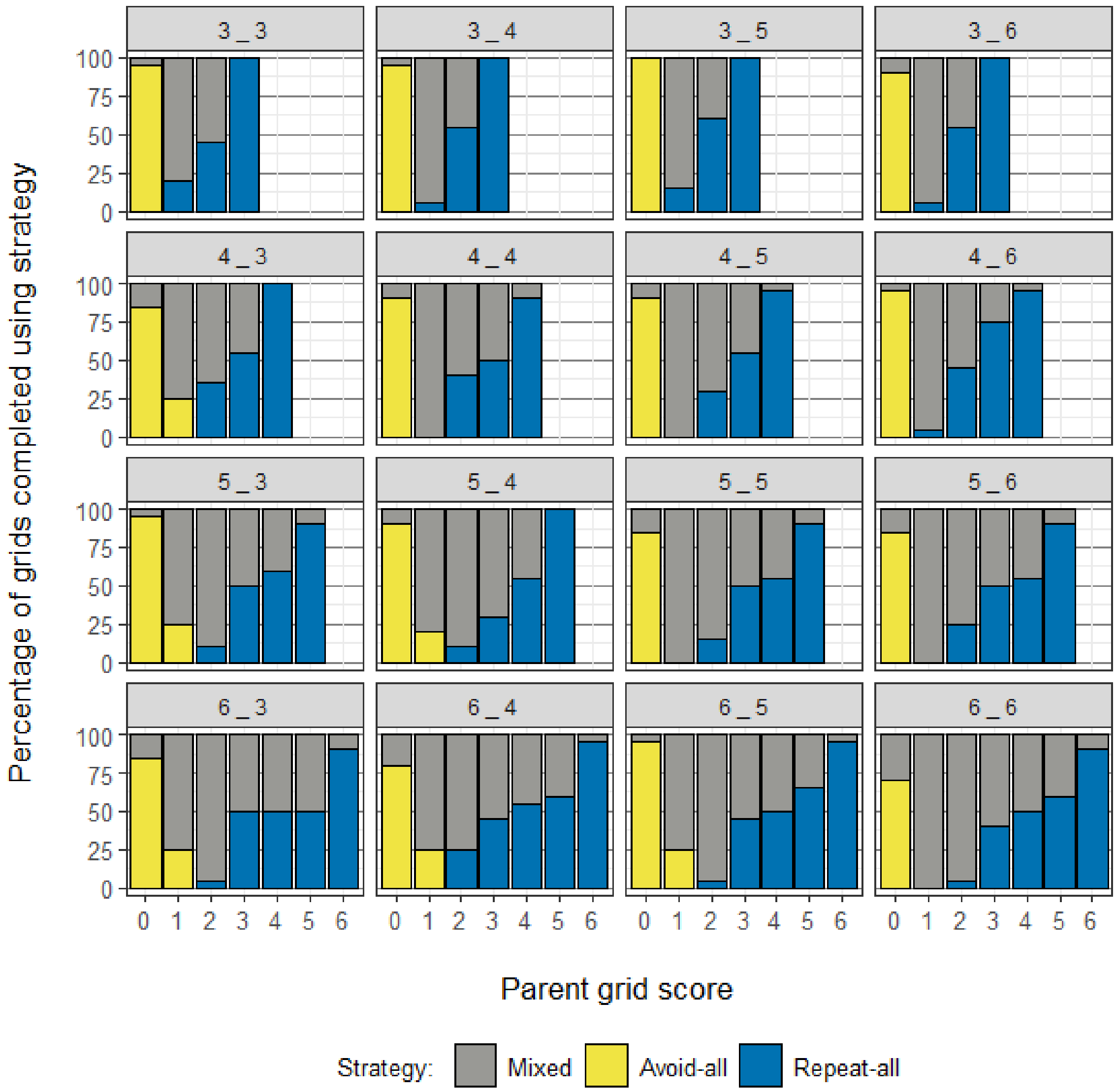
Figure 12 cont.



Note. The average score was taken from 500 separate simulations each run for 200 generations. Grid sizes are labeled as “number of rows” _ “number of stimuli on each row”. Red horizontal lines mark the maximum score for the grid size shown, the dark blue curve shows the simulated data, and the orange curve shows the ‘*low-risk performance strategy*’ for that grid, which was calculated by simulating all chain positions to perfectly apply a conditional ‘*repeat all*’ / ‘*avoid all*’ rule.

Figure 13

Percentage of participants that opted for each strategy for each size of landscape and parent score in the Low-precision condition.



Note. Where '*repeat-all'* and '*avoid-all'* strategies are shown, this was the lowest-risk strategy to use. There were no cases where a participant employed an '*avoid-all'* strategy where a '*repeat-all'* strategy was of the lowest risk, and vice versa. Grid sizes are labeled as "number of rows" _ "number of stimuli on each row".

2.4.3 Risk behaviour

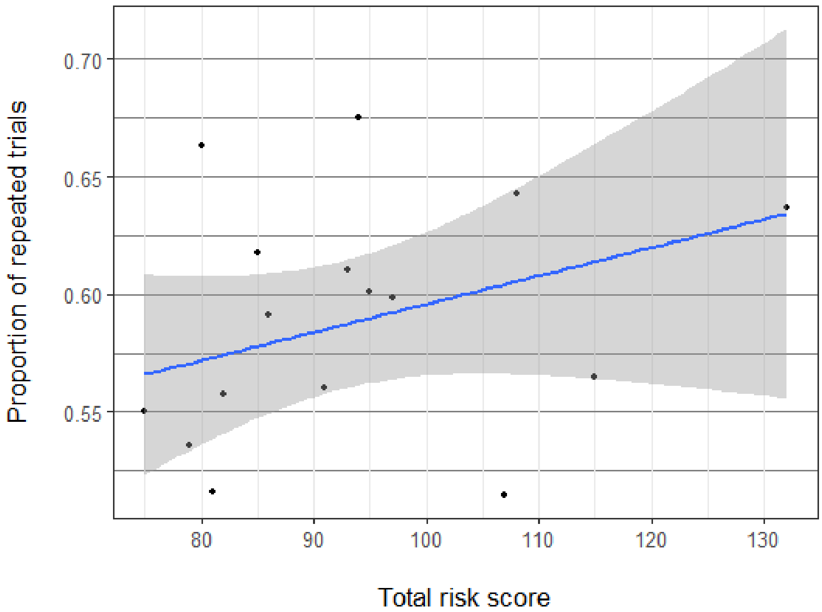
Statistical analysis was performed using R (R Core Team, 2020). P-values of < .05 were taken as statistically significant.

Risk behaviour was assessed using the total score from participants' assessment of their likely behaviour in 40 scenarios, each using a five-point rating scale. Higher numbers indicated a higher likelihood of undertaking a risk. Risk scores ranged from 75 to 132 (mean = 92.5, median = 88.5, n = 20).

We found no correlation between risk behaviour and the proportion of repeated trials, $r(18) = 0.304$, $p = .193$. (Figure 14) Because of this, we opted not to use this result to bolster any other findings.

Figure 14

Correlation graph showing the lack of correlation between the proportion of low-risk searches and the total risk score.



Note. Error areas represent +/- 1 standard error.

2.5 Discussion

The logistical challenges associated with using multi-generational transmission chain designs have severely limited the possibilities for investigating the mechanisms supporting distinctively human cumulative culture. In addressing these challenges, we provided an experimental demonstration of the PFR paradigm, introduced by Caldwell et al. (2020). As well as allowing for fully remote testing, this paradigm has the potential to reduce the number of participants required for generational designs, as well as the requirements for coordinating the participation of multiple individuals, by simulating repeated transmission from individuals' responses. Such a paradigm could open up the further possibility of accurately comparing multiple transmission conditions, age groups and species. in the current study, this was made possible by presenting each participant with a wide range of vicarious information representing many generations, as opposed to only presenting vicarious information from the predecessor in the same chain. This allowed assessment of the expected ratchet effect (i.e., scores increasing with transmission). In addition, this allowed for a more intricate assessment of repeating behaviour (i.e., how much of the model attempt the participant repeats, and deviates from).

Repeating Behaviour

We predicted that our High-precision condition would successfully satisfy the PFR criteria outlined by Caldwell et al. (2020). Specifically, we expected that participants would demonstrate the ability to significantly outperform a model demonstration, over multiple generations of transmission. By using this strategy, participants were able to match the parent model's performance for all rewarded selections and increase the chance of outperforming the parent model by deviating from selections that evidently resulted in no reward.

This result is perhaps unsurprising, given that a previous study using a similar task paradigm found that young children reliably copy successful demonstrations and shift from unrewarded demonstrations (Wilks et al., in press). While this developmental study could demonstrate only limited evidence of increase over '*generations*', it did provide support for the idea that capacities supporting cumulative culture develop in early childhood.

We had no particular expectations about the results in the Low-precision condition, however we did know that there was much less scope for individual participants to adopt a rule that ensured accumulation of beneficial information in most cases. The only rule that would guarantee no loss of information from the parent model (the '*low-risk strategy*') was to repeat all of the model selections if all, or even *some,* rewarded selections had been made by the parent model, however this restricted accumulation. The only situation in which participants should have risked deviating from the exact selections made by the parent model was in circumstances where the parent model scored below chance, and so scoring lower was unlikely with an avoid-all strategy. This strategy did, however, restrict the potential for achieving a perfect score to limited circumstances where avoiding all of the parent model selections resulted in selection of all targets by chance, which could only happen when the parent model's score was zero. Furthermore, selection of all of the targets in this circumstance was relatively unlikely, especially for larger grid sizes, where there was a decreased chance of selecting targets due to an increased number of stimuli per row, and an increased number of targets to uncover (more rows). In some cases, particularly in circumstances where only some of the parent selections were rewarded, real participant behaviour deviated from this low-risk strategy, and used the mixed strategy, whereby some parent selections were repeated, and others were avoided. Of course, because there was no indication as to which selections were rewarded, the strategy was applied somewhat arbitrarily, and on average this strategy could never outperform the low-risk repeat-all/avoid-all strategy as perfectly applied. However, adopting this strategy did open up the possibility of selecting all of the rewards by chance, although again this was less likely to occur in larger grid sizes.

Simulations of multi-generational transmission

Simulating repeated transmission demonstrated the effectiveness of this paradigm for assessing the potential for cumulative culture in two contrasting social learning contexts:

While the High-precision condition does appear to provide good evidence that human adults are capable of cumulative culture because of reasoning and insightful strategy, our simulations of the Low-precision condition suggest that these capacities are not completely necessary for accumulation to occur in all settings.

In simulating the behavioural data of the Low-precision condition, we found that individuals' use of a mixed strategy meant that transmission chains had the *potential* to reach the maximum score faster than the low-risk strategy, and in some grid sizes, participants were able to outperform a computer-generated parent model which was programmed to follow this low-risk strategy. Although individual participants adopting the mixed strategy do risk performing more poorly than the low-risk strategy, this increased their opportunities to select all of the rewarded stimuli by chance. Since even in real participant behaviour, a perfect parent model score usually resulted in a repeat-all response, these maximum scores – once achieved – were then relatively well-preserved within the chain simulations.

The results of this study successfully demonstrate that the PFR methodology is suitable for identifying cumulative culture, and for comparing conditions under which cumulative culture may be restricted in adult humans. As shown here, the vicarious information given to participants before they complete the task themselves does not necessarily need to be derived from a real participant sample. In this case, all vicarious information was artificially generated, and each participant saw every possible permutation of landscape size and score on each. This greatly reduced the requirement for a large participant pool, while providing results analogous to full chain studies. Using computer generated information also eliminated the need for coordination of participants, as no information transfer occurred.

The potential scope of the paradigm is broad. Even in applying close adaptions of the current task, such as to include multiple conditions or comparisons between species, it could contribute to answering a lot of open questions about cumulative culture. In particular, the method could be used to tackle those that are currently unanswered due to similar logistical challenges with implementation.

Methodological limitations

Of course, this paradigm is not without its limitations, and it would not be appropriate to claim that it would solve all the logistical problems associated with cumulative culture studies. The task itself is extremely abstract, with a requirement for participants to engage in behaviour that is very far removed from any real-world scenario typically associated with the phenomenon of cumulative culture. However, the flexible nature of the PFR paradigm means it may be possible to reduce the impact of this sacrifice in ecological validity by both altering the context of the task, and increasing its open-endedness. In this way, it might be possible to better approximate the challenges, and complexity of potential outcomes, of real-world social learning.

Conclusions

The current study demonstrated that cumulative culture can be successfully captured using adult participants in the PFR paradigm. This was evident from both individuals' repeating behaviour and simulated repeated transmission over generations. This paradigm therefore provides an effective alternative method for assessing the potential for cumulative cultural evolution in human and non-human populations. By making inferences about the outcome of repeated transmission based on individual performances, we reduce the requirement for complicated coordination of participants and large sample sizes. As a result, using this paradigm could open up the possibility of investigating questions about cumulative culture which were previously precluded as a result of the logistical challenges involved in implementing a traditional transmission design This method could therefore create opportunities to investigate the capacities required for cumulative culture in adult samples, as well as the developmental emergence of these capacities in children. Furthermore, it may open up possibilities to examine differences in cumulative culture between humans and non-humans, potentially only requiring minor adaptions, meaning the resulting data would be very comparable. Given the volume of research that has already been undertaken using established methods, we are optimistic that this paradigm will contribute greatly to expanding the field.

Chapter 3: Intentional information sharing promotes cumulative culture relative to inadvertent behavioural cues: an experimental demonstration

3.1 Abstract

Humans frequently communicate with intent to inform. This propensity, which is likely to be unique to humans, may contribute to the similarly distinctive capacity for cumulative culture. Here we study human adults, using an experimental transmission chain design, to investigate how intentional information sharing could promote the accumulation of beneficial information, relative to transmission via inadvertent behavioural cues. Participants completed a landscape-searching task, scoring points by finding hidden targets. Information was transmitted between participants such that receivers were informed of the results of part or all of their predecessor's search attempt. There were three information conditions: Intentional, Inadvertent, and Full. In the Intentional and Inadvertent conditions, a small subset of the search was transmitted from cultural parents to cultural offspring, either selected by the information producer themselves for informativeness (Intentional), or a random sample (Inadvertent). Scores increased over learner generations across all conditions, but were higher in the Intentional condition compared with the Inadvertent condition. Furthermore, scores in the Intentional condition were comparable to those in the Full condition, despite participants being provided with a more limited sample. We conclude that intentional information sharing can compensate for loss of expertise that would otherwise occur as a consequence of transmission bottlenecks (e.g., limited learning time).

3.2 Introduction

Cumulative cultural evolution (Caldwell & Millen, 2008b; Mesoudi & Thornton, 2018) is widely recognised to be both pervasive and ubiquitous in human societies; techniques, technologies and practices are not only socially transmitted, but also become further developed over time, through refinements, modifications and extensions. This process is therefore responsible for one of the most interesting and noteworthy properties of human culture, which is that later generations of particular populations are generally able to benefit from valuable knowledge and resources that were unavailable to their predecessors (e.g., Caldwell (2018)). In contrast, evidence of equivalent phenomena appears rare in non-humans (e.g., Dean et al. (2014); although see Sasaki & Biro (2017); and Jesmer et al. (2018); for some noteworthy exceptions).

Most accounts of the distinctiveness of human cumulative culture have focussed on the role of cognitive mechanisms that also appear to be restricted to humans (e.g., Boyd et al. (2011); Dean et al. (2014); Hill et al. (2009)). There are numerous theoretical accounts which propose compelling logical rationales linking such individual-level cognitive abilities to the potential for cultural accumulation at the population-level (e.g., Boyd & Richerson, (1996); Tennie et al. (2009); Tomasello et al. (2005)). However, there is still relatively little experimental evidence investigating how the identified factors actually influence group-level outcomes, and under what circumstances.

Here we report an experiment investigating the accumulation of beneficial information over generations, comparing the effects of inadvertent social information with intentional information sharing. This comparison may be particularly relevant to understanding differences between human and non-human cultural evolution, as humans readily engage in intentional knowledge sharing, whereas the vast majority of social information use in non-humans depends on inadvertent social information. Inadvertent social information is used to describe information available as a consequence of others' efficient performance of their activities (Danchin et al., 2004). As such, it is neither selected for, nor intended to, perform any communicative function for the cultural offspring. In spite of this, cultural offspring may nonetheless be able to use such '*public information'* to their advantage. For example, others' foraging success may provide valuable information about patch quality (Smith et al., 1999; Templeton & Giraldeau, 1996). Learning from such inadvertent social information even appears to be the main mechanism supporting cases of relatively enduring behavioural traditions in non-humans. For example, chimpanzees learn the foraging and social behaviours particular to their social group (e.g., Whiten et al. (1999)), but there is little or no evidence of experienced individuals modifying their behaviour in ways that would facilitate others' learning (e.g., Thornton & Raihani (2008); Whiten et al. (2003)). The chimpanzees' learning therefore appears to occur primarily through exposure to information which is available purely incidentally as a result of others' activity.

It should nonetheless be emphasised that some non-human behaviours do perform a teaching function (e.g., see review in Thornton & Raihani (2008)). However, these examples appear to be the exception, rather than the rule. Furthermore, although such cases cannot be classified as involving inadvertent social information (since the information producer plays an active role in facilitating transmission, (e.g., Hoppitt et al. (2008)), they nonetheless do not reflect intentional information sharing. Examples of teaching in non-humans appear to be fairly narrowly-focused adaptations that facilitate learning in the very specific context of particular species-typical behaviours (e.g., prey provisioning in meerkats, Thornton & MaCuliffe (2006); and tandem-running in ants, Franks & Richardson (2006). It is unlikely that the animals in question have any insight into the fact that their behaviour impacts on others' knowledge or skill level. As Premack (2007) has previously noted, non-human teaching reflects adaptations that have a single target, whereas human teaching is a "domain-general competence with indeterminately many targets" (p13862).

The domain-generality of human teaching likely arises as a consequence of an individual's recognition that their actions can potentially benefit others' learning, along with some understanding of how this might be achieved, thus allowing for intentional knowledge sharing. In contrast it is widely agreed that non-human teaching probably does not involve this kind of understanding, due to disparities in capacities for mental state understanding (e.g., Kline (2015); Thornton & Raihani (2008)).

It is perhaps unsurprising then that such distinctively human teaching has been identified by a number of authors as potentially a key mechanism that could explain the evolutionary anomaly of human cumulative culture (Caldwell et al., 2018; Dean et al., 2012; Kline, 2015; Laland, 2017). Indeed, Caldwell et al. (2018) have noted that human intentional knowledge sharing ensures that the kind of novel variants inherent to cumulative culture can be taught to others, despite the absence of any opportunity for adaptations to evolve which are specific to these traits.

Despite the relative abundance of theoretical arguments linking them, there is currently little evidence demonstrating that intentional knowledge sharing can indeed facilitate the accumulation of beneficial information, over and above the use of inadvertent social information. Experimental research investigating cumulative culture under laboratory conditions (with participants organised into transmission chains or microsocieties) have sometimes involved manipulations of learning conditions (e.g., Caldwell & Millen (2009); Zwirner & Thornton (2015)), although also see Morgan et al. (2015)). However, to date these studies have found little evidence of benefits from intentional knowledge sharing. Although neither study was directly focussed on the question of comparing intentional with inadvertent social information, both Caldwell & Millen (2009) and Zwirner & Thornton (2015) included conditions involving teaching by verbal communication, and compared these with conditions in which participants were exposed to only the finished products left by their predecessors. These conditions could therefore in principle be re-cast as involving intentional information sharing and inadvertent social information respectively. However, the differences between these conditions extended well beyond the fact that the information was either intentionally or inadvertently shared, with participants also exposed to a much greater *quantity* of information in the teaching conditions, compared with the products-only conditions. Furthermore, despite this, task scores were found to increase over learner generations even in the products-only conditions of both studies, with very little evidence of additional advantages from teaching.

In the current study, we aimed to directly compare conditions which differ only with respect to whether the transmitted information has been intentionally shared or is acquired incidentally from the information producer's activity. In addition, the goal of this study extends beyond the simple question of whether it is possible to identify an advantage of one transmission condition over another, but to instead identify conditions under which this might be the case. As Caldwell (2018) has previously argued, laboratory studies of cultural evolution must be interpreted with caution in relation to identifying prerequisites and constraints on cumulative culture. Whilst the studies by Caldwell & Millen (2009) and Zwirner & Thornton (2015) demonstrate that intentional knowledge sharing is not a strict prerequisite for cumulative culture, it is possible that there may be circumstances under which it offers significant advantages.

Here we test the hypothesis that when transmission occurs through a learning bottleneck (i.e., only limited information is transmitted), intentional knowledge sharing will allow for the accumulation of beneficial information, as compared with inadvertent social information, due to the potential for highlighting information likely to hold the greatest value for the learner.

Participants took part in a task which required them to search for hidden targets within landscapes which were subdivided using a grid. Points on the grid could be searched one at a time, and participants scored points by finding the targets, using a limited lifespan of a finite number of unsuccessful search attempts. Information transmitted between participants included the location of grid points searched by the cultural parent, but also whether this was a hit (part of a target) or a miss.

We ran 15 chains each composed of ten participants, with five chains in each of three conditions: Full, Intentional and Inadvertent information. In the Full information condition, cultural offspring were shown all of the search outcome obtained by the cultural parent. In the Intentional information condition, each participant, in their role as cultural parent after completing their grid search in the role of cultural offspring, was asked to select a subset of the hits and misses they had revealed during the search stage (three grid squares) for transmission to the next cultural offspring. In the Inadvertent information condition, an equivalent subset (three grid squares) was randomly selected from the outcome revealed during each participant’s search, and this formed the information transmitted to the next participant in the chain (see Methods for full details).

We predicted that participants’ scores would be higher in the Intentional information condition, compared with the Inadvertent information condition, and that this difference would be more apparent in later generations, compared with earlier ones (indicative of the benefits arising as a result of the accumulation of information enabled by social transmission). Scores in the Full information condition were expected to be highest of all, with these providing an indication of the upper limit of the benefits that could accumulate as a consequence of social learning in this task context. Our key predictions were pre-registered: https://osf.io/jszvc/

3.3 Methods

3.3.1 Participants

One hundred and fifty-nine adults aged between 17 and 62 years (Median = 20 years, 103 females) were recruited in total, with data from nine of these participants being excluded from the study due to experimenter or technical error (in accordance with the exclusion criteria specified in the pre-registration). The final sample consisted of 150 adult participants aged between 17 and 62 years (Median = 20, 100 females). Of the final sample, 133 participants were recruited from the University of Stirling via an online recruitment database, through which students can earn tokens required for course completion by participating in research studies, and a further 17 participants were recruited from Glasgow Science Centre (www.glasgowsciencecentre.org/) on a voluntary basis.

Fifty participants were assigned to each of the three conditions, with ten participants in each transmission chain, and five complete chains in each condition. Participants were assigned to a condition randomly, and chain position was constrained by the current state of completion of the chain to which they were recruited. Participants whose data was excluded were replaced in that chain by a new participant, prior to further recruitment to that particular chain. The study was approved by the University of Stirling General University Ethics Panel (approval reference number: GUEP 295), and every participant gave written consent before the experiment took place.

3.3.2 The experimental task

The task completed by the participants was a landscape searching challenge loosely based on the game '*Battleships'*. The task was run using PsychoPy 1.84.2 (Peirce et al., 2019), on either a desktop computer or a Microsoft Office tablet running Windows 10.

Participants were presented with a series of ten 16x16 grids, in each of which three 3x3 targets (ships) could be found. Targets were randomly placed, and consisted of nine grid points (See Figure 15). Participants could search the grid for targets by clicking on (desktop), or touching (tablet), any of the grid squares. When part of a target was found, these were scored as '*hits'*, and selection of any of the non-target squares in the 16x16 grid were scored as '*misses'*. The participant's primary goal was to maximise their score by finding as many hits as possible, across the ten grids.

Participants used up one of their limited search attempts for every grid square they selected which did not contain part of a target. In contrast, when participants selected a target grid square, the number of search attempts did not reduce, and they were instead awarded one point. This scoring system was intended to capture the relative payoffs typically associated with real world subsistence activity, with fruitless search always costly, and success always profitable. A maximum score of 27 points was available in each grid (nine for every full target), and participants were allotted nine search attempts in which to find them. This meant that the score for each grid varied depending on how many hits and misses were found. For example, if all 27 hits (i.e., three complete targets) were found without any error, the score for that grid would be 27, with nine search attempts remaining and exactly 27 grid squares revealed. If no targets were found, the score for that grid would be zero, with all search attempts used and only nine (miss) grid squares revealed.

The first participant in each chain received no information about the contents of any of the grid squares, other than that generated by their own search. All other participants in the chain received some information about their cultural parent's performance for each grid.

The same pre-specified reward landscape was kept consistent for each grid completed by a particular chain, such that the results of the cultural parent's search held true for the cultural offspring. Depending on the condition, participants saw some or all of the search outcome obtained by their cultural parent (including whether each revealed grid square was a hit or a miss). The nature of the information varied by condition as detailed below:

Full Information (control): With the exception of the first in each chain, each participant was shown all of the search outcome obtained by their cultural parent. Thus, cultural offspring received information about the value of between nine and 36 grid squares.

Intentional Information: At the end of their search, each participant in this condition was asked to select a subset of the hits and misses they had revealed during the search stage (three grid squares). Thus, the information that each participant received comprised the value of these three squares, as selected by their cultural parent in the chain.

Inadvertent Information: A randomly selected subset (three) of the grid squares revealed during each participant's search were passed on to the next participant in the chain. Thus, as in the Intentional condition, participants in this condition received information about the value of three grid squares. This condition was therefore intended to be analogous to using '*public information*' (e.g., Danchin et al., 2004), available as a by-product of another's activity, as long as the observer happens to be in the vicinity of the cultural parent.

3.3.3 Procedure

Participants gave written consent before being pseudo-randomly assigned a condition and chain position according to the constraints detailed in the Participants section (3.3.1).

Participants were provided with thorough instructions on screen before beginning the experiment. These included picture examples of an unselected grid square, a selected hit, a selected miss and a full 3x3 target. Participants were also given instructions about the constraints on their search activity for any given grid, i.e., that the program would advance to a new grid when either all nine search attempts had been used up, or when all 27 hits (i.e., three complete targets) had been selected (see description of experimental task above).

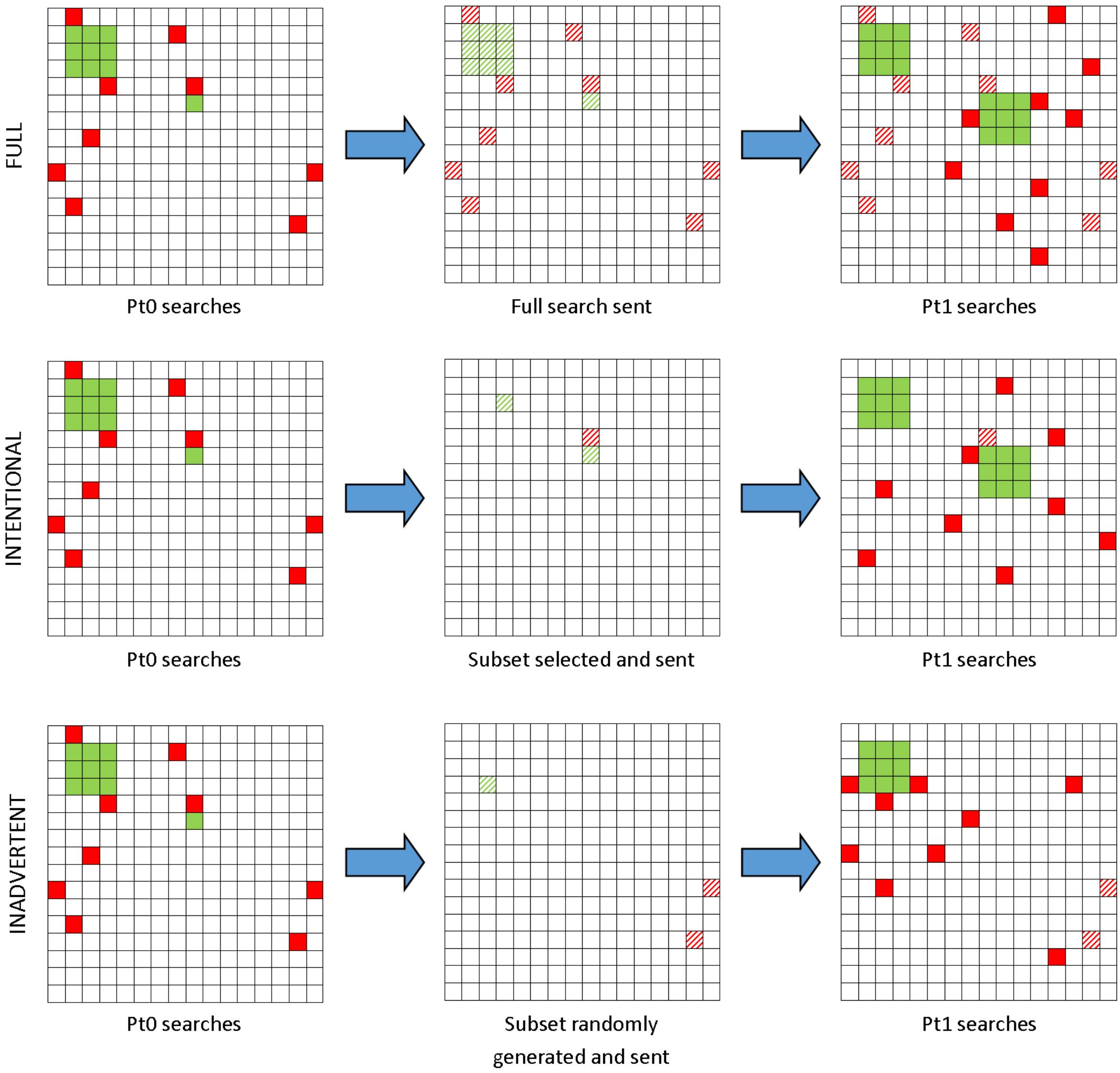
If a participant was in generation 1 (i.e., the first participant in a chain of ten), then they received no information about the location of hits or misses. Participants in all other chain positions received some information about the grid, as detailed in the description of the experimental task. In all conditions, participants were told that they would receive *some information* from a previous participant's performance, and that they could use this information however they wished. Details about the source of the information was not explicitly disclosed to participants.

After participants had completed their own search, in the Intentional condition, they were also required to make decisions about the information to be transmitted to the next participant in the chain. Participants in this condition were therefore made aware that the information was to be sent to a future participant who would be presented with the same task they had just completed. Following each grid search, they were asked to select three of the grid squares from the selections that they had made for any given grid.

After all ten grids had been completed, participants were shown their final score (total hits selected across all ten grids). Participants were then fully debriefed regarding the aims and hypotheses of the experiment, and given the opportunity to ask any questions.

Figure 15

Demonstration of information transmission between two generations.



Note. In this demonstration, filled green grid squares indicate a hit, filled red grid squares indicate a miss and striped grid squares indicate information that was transmitted from the cultural parent but not subsequently selected by the cultural offspring. Blue arrows represent the direction of transmission.

3.4 Results

All generalised linear mixed models were performed using R (R Core Team, 2020) and lme4 (Bates et al., 2014). Graphs were created using *ggplot 2* performed on R Studio (RStudio Team, 2020), supported by *R version 4.0.2*. The dependant variable in each analysis was score or search attempts remaining (both numeric). Chain position was centred every time. Fixed effects are specified before each model output.

Non-convergent and singular fit models were addressed by removing random effects, however models with the maximum random effects structure (Barr et al., 2013) were considered first. Post-hoc comparisons were done using the *emmeans* package (Lenth et al., 2019). Tests were two-tailed, and *p*-values of < .05 were taken as statistically significant. Emmeans using the Tukey adjustment was used for multiple comparisons, and Emtrends was used to assess interactions between slopes when significant interactions were found in the model.

3.4.1 Mean scores

Mean scores were calculated using each participant's sum scores across all 10 completed grids. The maximum score possible for participants to achieve was 270, that is, 27 points for each of the ten grids searched.

First, we examined the scores at chain position 1 using condition as the independent variable, and chain included as a random variable. The Intentional condition was used as the intercept. We found no significant differences in scores at chain position 1 between the Intentional and Full conditions (b = -20.8, SE = 18.5, $t(8)$ = -1.123, p = .528), between the Intentional and Inadvertent conditions (b = -4.4, SE = 18.5, $t(8)$ = -0.238, p = .970) or between the Full and Inadvertent conditions (b = 16.4, SE = 18.5, $t(8)$ = 0.885, p = .664). (Figure 16a) Any significant differences at the beginning of the chains might have cast doubt on the source of any differences that we found in later generations of the chains (i.e., potentially suggesting that these might not be a result of differences in the degree of accumulation of beneficial information).

In this model, we included condition and chain position (centred) as independent variables, chain was included as a random variable. The Intentional condition was used as the intercept.

Looking at all three conditions together, a significant main effect was found for chain position (b = 116.39, SE = 28.34, t(149) = 4.12, p < .001). There were also main effects of condition. Sum score was significantly higher in the Intentional condition when compared to the Inadvertent condition (b = 40.6, SE = 11.5, t(140) = 3.529, p = .002), and significantly higher in the Full condition when compared to the Inadvertent condition (b = 63.2, SE = 11.5, t(140) = 5.491, p < .001). No significant difference was found between the Full and Intentional conditions (b = 22.6, SE = 11.5, t(140) = 1.96, p = .126).

We also tested for interactions between condition and chain position for the sum score measure. This would tell us whether any differences between conditions increased over generations, in line with our predictions. The comparisons between conditions, analysing interactions involving chain position, were as follows: Full and Intentional conditions (b = -21.7, SE = 40.1, t(140) = -0.541, p = .852); Full and Inadvertent conditions (b = 50, SE = 40.1, t(140) = 1.248, p = .427); and Intentional and Inadvertent conditions (b = 71.7, SE = 40.1, t(140) = 1.789, p = .177).

3.4.2 Exploratory analysis of differences between conditions using only generations 1-6

In the event of finding support for our predictions regarding overall differences in sum score between conditions (which we did indeed find), we had expected also to find these differences accompanied by corresponding interactions involving chain position. If the differences between the conditions were more apparent in later generations, this would be consistent with the idea that these differences were attributable to differences in the degree to which beneficial information was accumulating over generations. However, as reported above, we did not find any significant interactions between condition and chain position, for the measure of sum score. We suspected that the absence of interactions between conditions could have been attributed to scores reaching a plateau relatively early in the chains and improving no further beyond this point. In our planned analysis (which considered all 10 generations) the plateau effect in later generations could have potentially obscured the effect any differences in rates of increase between the conditions in earlier generations. In Figure 16b it can be seen that the plateau begins around chain position 6 (with all conditions being similar in this respect, despite other differences). Indeed, there even is some indication that scores began to drop after this point. In the Discussion we will return to the issue of why task performance may have actively deteriorated in later generations, possibly as a direct consequence of the task becoming so trivial that this caused confusion. However, setting aside the issue of interpretation, here we simply wished to explore the possibility of finding interactions between conditions using only data generated prior to the plateau. We therefore re-ran the analysis reported above, with chain positions 7-10 omitted. (Figure 16c)

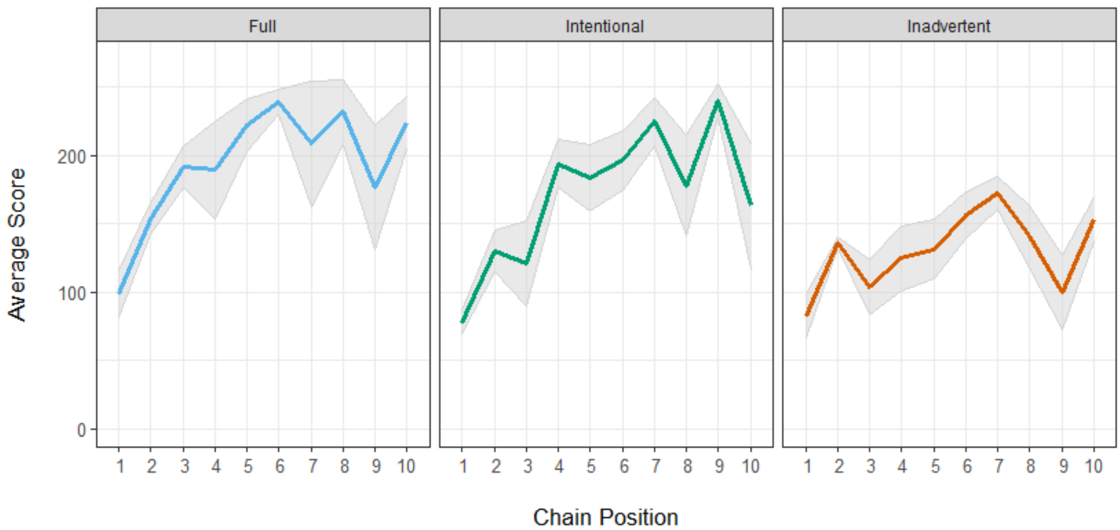
In this model, we included condition and chain position (centred) as independent variables, and chain is included as a random variable. The Intentional condition was used as the intercept.

Broadly speaking, the main effects and pairwise comparisons between conditions showed a similar pattern of results to the planned analysis. A significant main effect was found for chain position (b = 141.189, SE = 27.638, t(89) = 5.108, p < .001). Sum score was not significantly higher in the Intentional condition when compared to the Inadvertent condition (b = 27.9, SE = 11.1, t(80) = 2.511, p = .037. Scores in the Full condition were significantly higher compared with scores in the Intentional conditions (b = 31.8, SE = 11.1, t(80) = 2.858, p = .015), and significantly higher in the Full condition when compared to the Inadvertent condition (b = 59.7, SE = 11.1, t(80) = 5.369, p < .001). This comparison identified a predicted difference between these conditions in this analysis, which had not been found in the planned analysis reported above. This may have been attributable to the fact that any confusion caused by the excessive easiness of the task (mentioned previously as a possible explanation for the observed plateaus/peaks) would have had a dampening effect on the scores in later generations in the Full Information in particular.

Nonetheless, we were most interested in the question of whether some interaction effects involving chain position might be significant once any generations subject to plateau effects had been excluded. The comparisons between conditions, analysing interactions involving chain position, were as follows: Full and Intentional conditions (b = 13.9, SE = 39.1, $t(80)$ = 0.354, p = .933); Full and Inadvertent conditions (b = 90.2, SE = 39.1, $t(80)$ = 2.309, p = .060); and Intentional and Inadvertent conditions (b = 76.4, SE = 39.1, $t(80)$ = 1.954, p = .130). Therefore, no significant interactions were found, despite this being marginal between the Full and Inadvertent conditions. However, it is perhaps worth noting that in this analysis the identified trends corresponded more closely to the pattern of results that we had predicted. Also, the interaction for which we would have made the strongest predication, had a relatively low p-value.

Figure 16a

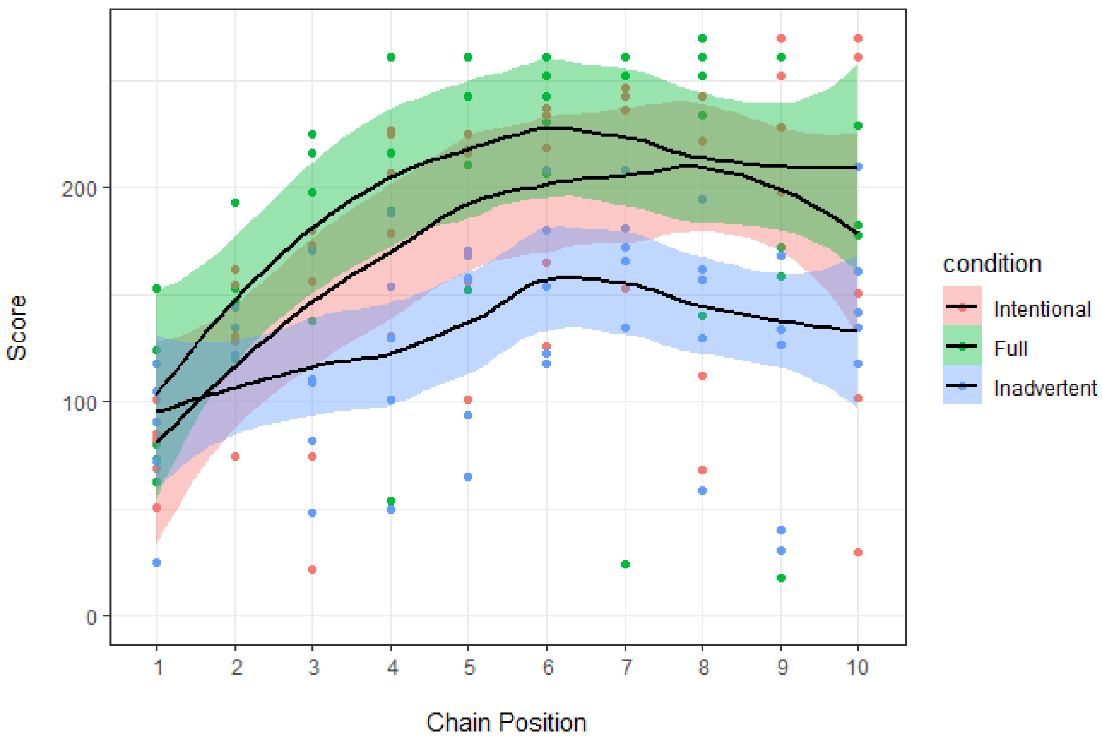
Mean score increase at each chain position, split by condition.



Note. Error areas represent +/- 1 standard deviation.

Figure 16b

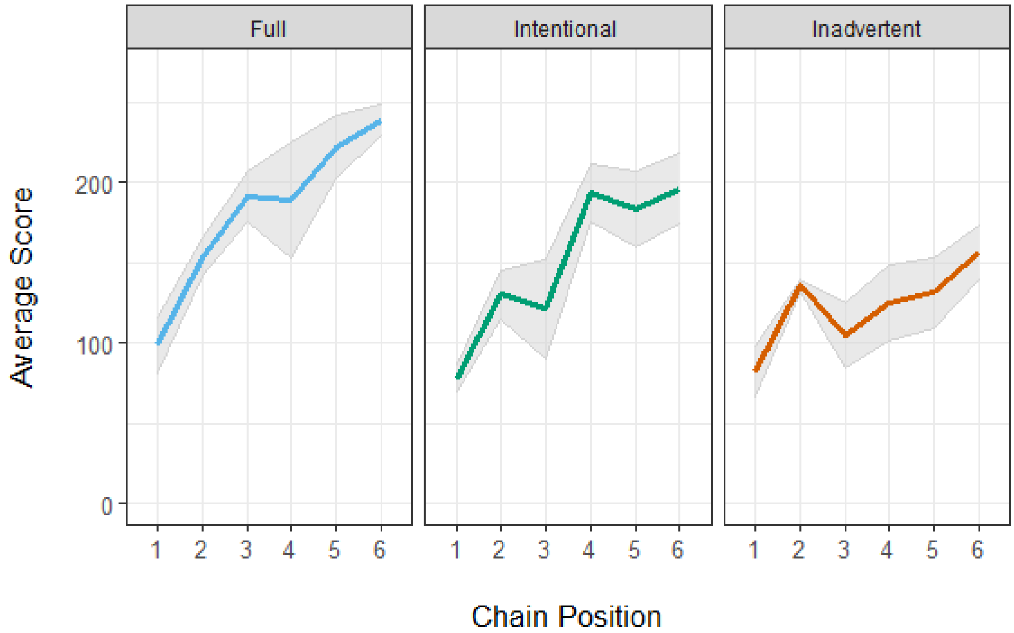
Curvilinear regression graph showing the plateau of scores in each condition.



Note. Data points have been jittered so that the number of points at each value is visible. Error areas represent +/- 1 Standard Error.

Figure 16c

Mean score increase from chain position 1 to chain position 6, split by condition.



Note. Error areas represent +/- 1 standard deviation.

3.4.3 Mean search attempts remaining

Mean search attempts remaining were calculated using each participant's sum scores and search attempts lost across all 10 completed grids. The maximum possible number of search attempts remaining was 90, that is, nine search attempts for each of the ten grids searched.

First, we examined the search attempts remaining at chain position 1 using condition as the independent variable, and chain included as a random variable. The Intentional condition was used as the intercept. We found no significant differences in search attempts remaining at chain position 1 between the Intentional and Full conditions (b = -0.4, SE = 0.86, $t(8)$ = -0.463, p = .890), between the Intentional and Inadvertent conditions (b = -1.4, SE = 0.86, $t(8)$ = -1.620, p = .292) or between the Full and Inadvertent conditions (b = -1.0, SE = 0.86, $t(8)$ = -1.157, p = .509) (Figure 17). As with score, this test provided reassurance that there was no pre-existing difference between the conditions prior to any information transmission.

In this model, we included condition and chain position (centred) as independent variables, chain was included as a random variable. The Intentional condition was used as the intercept.

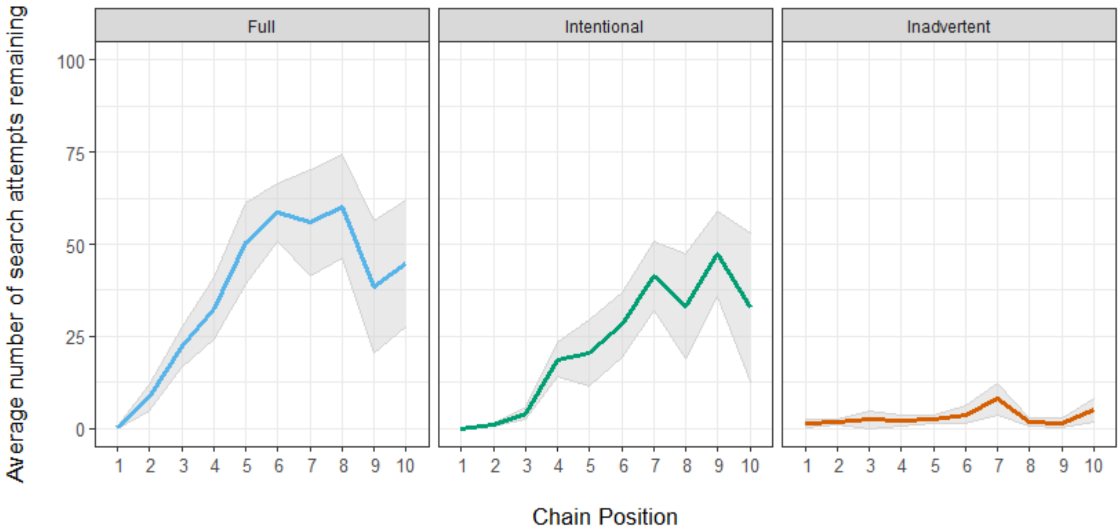
Looking at all three conditions together, a significant main effect was found for chain position (b = 50.97, SE = 9.64, $t(149)$ = 5.286, p < .001). There were also main effects of condition. The sum of the search attempts remaining was significantly higher in the Intentional condition when compared to the Inadvertent condition (b = 19.7, SE = 3.92, $t(140)$ = 5.035, p < .001), and significantly higher in the Full condition when compared to the Inadvertent condition (b = 34.1, SE = 3.92, $t(140)$ = 8.706, p < .001. There were also significantly more search attempts remaining in the Full condition compared with the Intentional condition (b = 14.4, SE = 3.92, $t(140)$ = 3.671, p = .001).

We also tested for interactions between condition and chain position for the search attempts remaining measure. This would tell us whether any differences between conditions decreased over generations, in line with our predictions. The comparisons between conditions, analysing interactions involving chain position, were as follows: Full and Inadvertent conditions (b = 50.36, SE = 13.60, $t(140)$ = 3.693, p < .001);

Intentional and Inadvertent conditions (b = 48.15, SE = 13.6, t(140) = 3.531, p = .002; and Full and Intentional conditions (b = 2.22, SE = 13.6, t(140) = 0.163, p = .985).

Figure 17

Mean search attempts remaining at each chain position for each condition, split by condition.



Note. Error areas represent +/- 1 standard deviation.

3.4.4 Survival analysis for the first point of completion for each grid

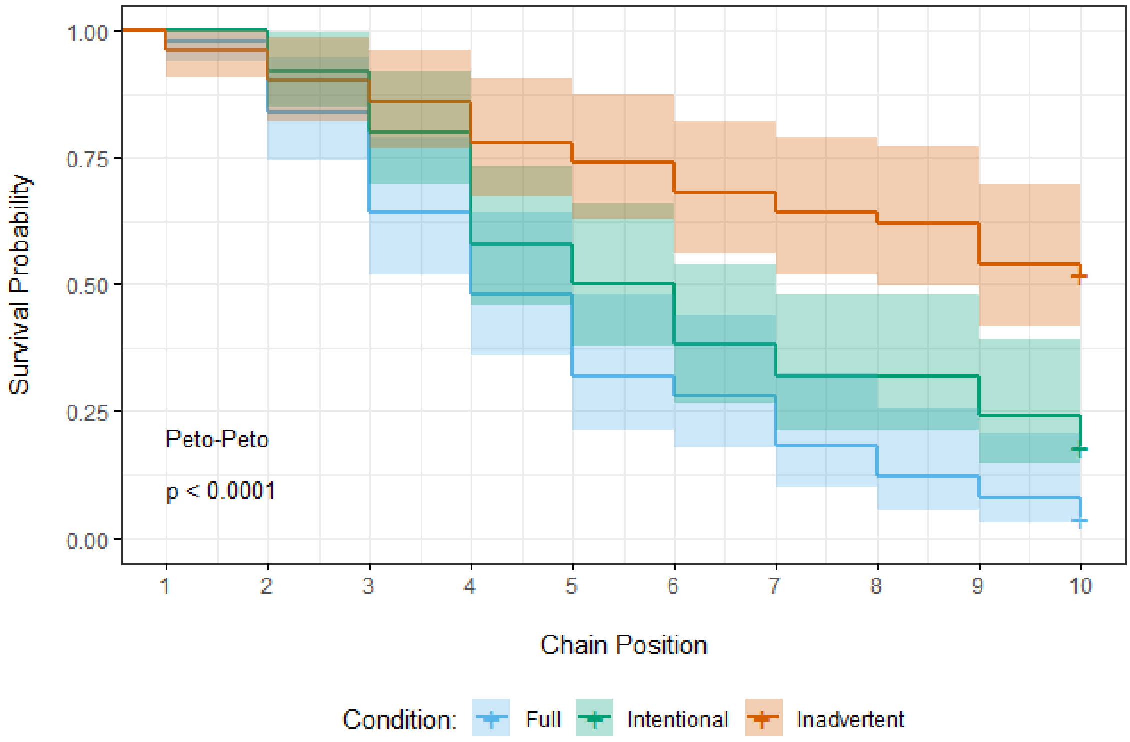
Kaplan-Meier survival analysis was used to construct a curve that displays the rate of first solution at each chain position, detailing the earliest point in the chain that all 3 full targets are uncovered (target maximisation). Stratified analysis using the Peto-Peto test was used to compare completion of grids in each condition – the Peto-Peto test was selected because it assumes proportional hazards are not constant This is primarily used when events (i.e., all 3 full targets are uncovered) are weighted depending on the percentage chance of estimated failure, for example, early events are allocated more weight because the probability of that event taking place is lower early in the chain (Karadeniz & Ercan, 2017). Cox proportional hazard regression analysis was used to identify which condition predicted the successful completion of grids, and estimate hazard ratios with 95% confidence intervals.

Figure 18a shows the survival curve for all grids completed by participants at each position in the chain. The stratified analysis showed significant differences in median completion time over conditions ($p < .001$).

Figure 18b shows the Cox proportional hazard regression model. Compared to the Intentional condition, the chance of completion was significantly higher in the Full condition (*HR* 1.632, 95%, *CI* 1.072 to 2.485, $p = .022$). The chance of completion in the Inadvertent condition was significantly lower than in the Intentional condition (*HR* 0.403, 95%, *CI* 0.243 to 0.669, $p < .001$).

Figure 18a

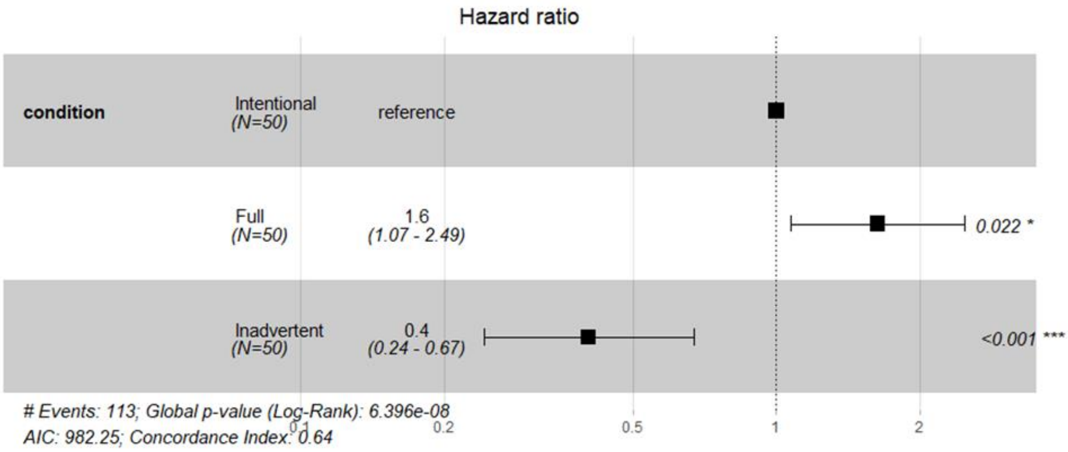
Survival curve for all grids completed by participants at each position in the chain



Note. Error areas represent +/- 1 Confidence Interval.

Figure 18b

Forest plot of hazard rations from Cox proportional hazard regression model, using the Intentional condition as a baseline.



Note. Error areas represent +/- 1 Confidence Interval.

3.4.5 Scores and search attempts remaining following information from a complete grid

This analysis compared conditions using chain as a random variable.

The data used for this analysis only included grids that followed a grid in which all three targets were found. (Full condition, n = 219; Intentional condition, n = 147; Inadvertent condition, n = 43).

Scores were found to be significantly higher in the Full condition (n = 219) compared to scores in the Intentional condition (n = 147) (b = 3.07, SE = 0.844, t(406) = 3.635, p < .001). Scores in the Inadvertent condition (n = 43) were found to be significantly lower than scores in the Full condition (n = 219) (b = -5.12, SE = 1.342, t(399) = -3.811, p < .001). Scores in the Intentional condition (n = 147) were not found to be significantly higher than scores in the Inadvertent condition (n = 43) (b = 2.05, SE = 1.423, t(374) = 1.440, p = .322). (Figure 19a)

Search attempts remaining in the Full condition (n = 219) were found to be significantly higher than search attempts remaining in the Intentional condition (n = 147) (b = 3.46, SE = 0.313, t(405) = 11.044, p < .001). Search attempts remaining in the Full condition (n = 219) were found to be significantly higher than search attempts remaining in the Inadvertent condition (n = 43) (b = 6.67, SE = 0.498, t(406) = 13.387, p < .001). Search attempts remaining in the Inadvertent condition (n = 43) were significantly fewer than search attempts remaining in the Intentional condition (n = 147) (b = -3.21, SE = 0.53, t(405) = -6.076, p < .001). (Figure 19b)

Figure 19a

Scores at the end of each grid search for participants who were given information from a previous grid where all 27 target points had been found.

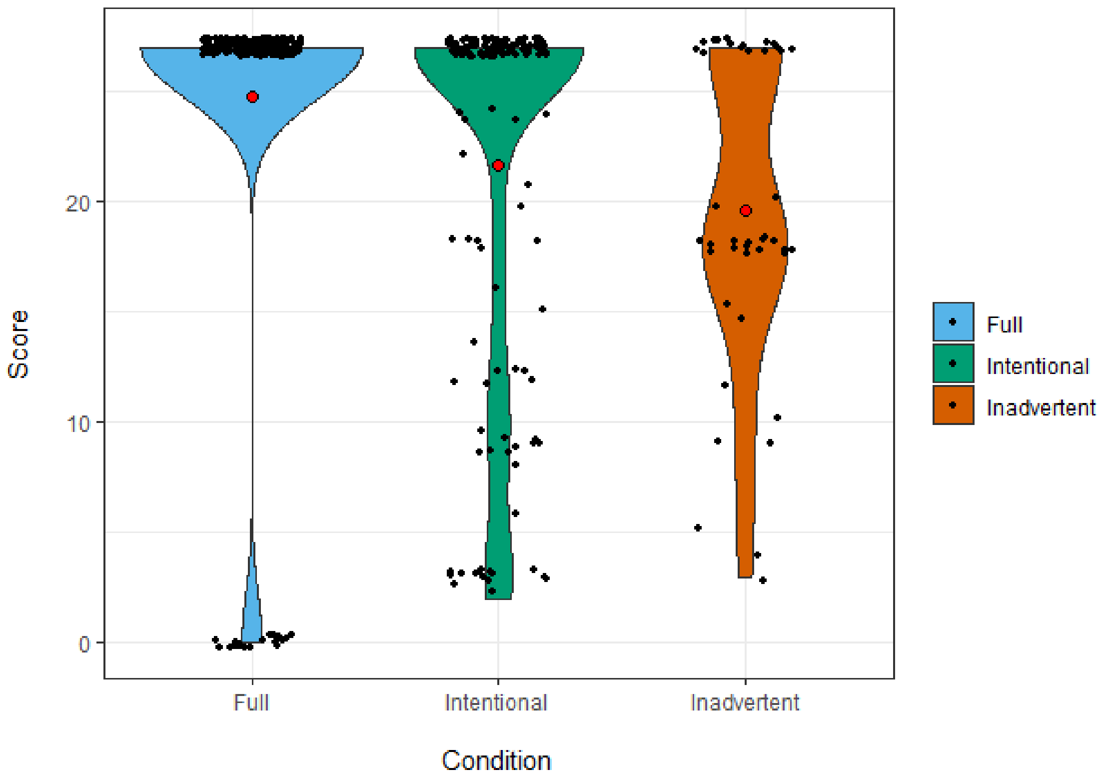
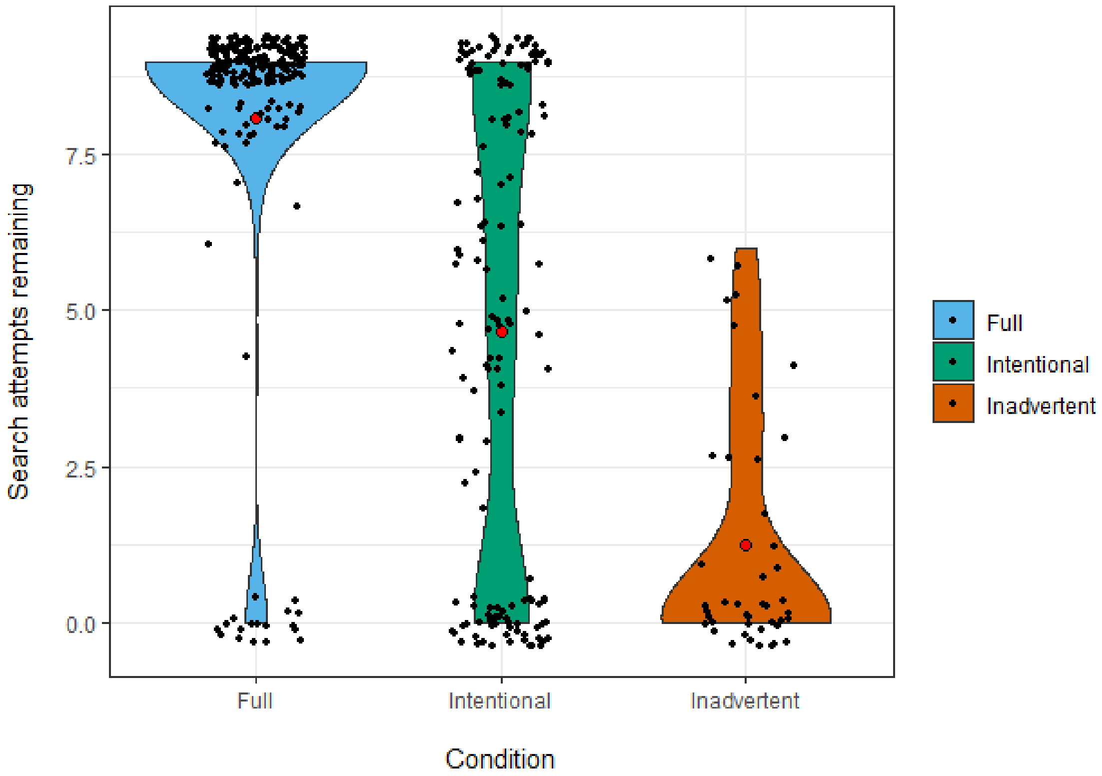


Figure 19b

Search attempts remaining at the end of each grid search for participants who were given in formation from a previous grid where all 27 target points had been found.



3.4.6 Proportion of hits transmitted

This analysis only included data from the Intentional condition.

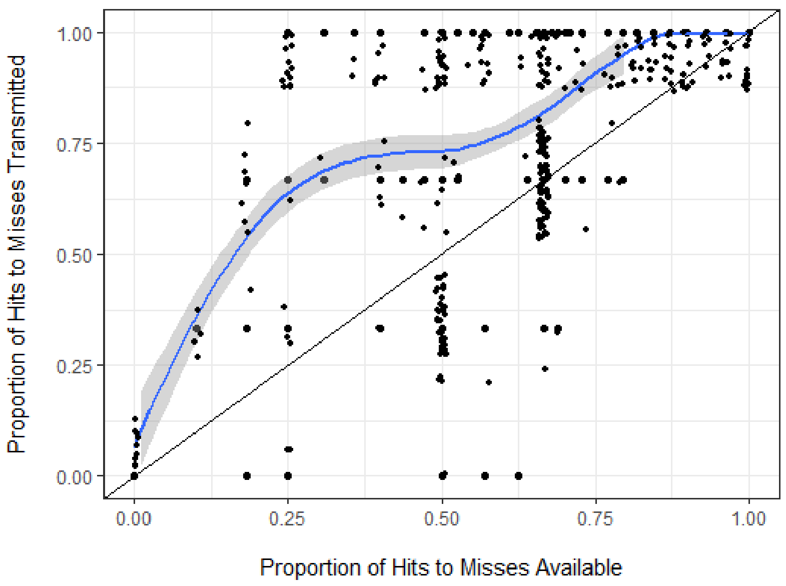
A paired sample t-test was used to compare the proportion of hits to misses sent by participants, to the total proportion of hits to misses they had found in their own search (thus, the proportion of hits to misses potentially available to send).

On average, participants had 16.67 hits available and 7.04 misses available per grid. The average hits available to total grid squares available was 63.66%.

The proportion of hits to misses sent was significantly higher than the proportion of hits to misses available ($t(499) = 15.89$, $p < .001$). This indicates that participants selected a higher proportion of hits than would be expected according to a random selection of the grid points available to send.

Figure 20

Proportion of hits to misses available compared with the proportion of hits to misses transmitted in the Intentional condition only.



Note. Data points have been jittered so that the number of points at each value is visible. Error areas represent +/- 1 Standard Error.

3.4.7 *Target segment positions transmitted*

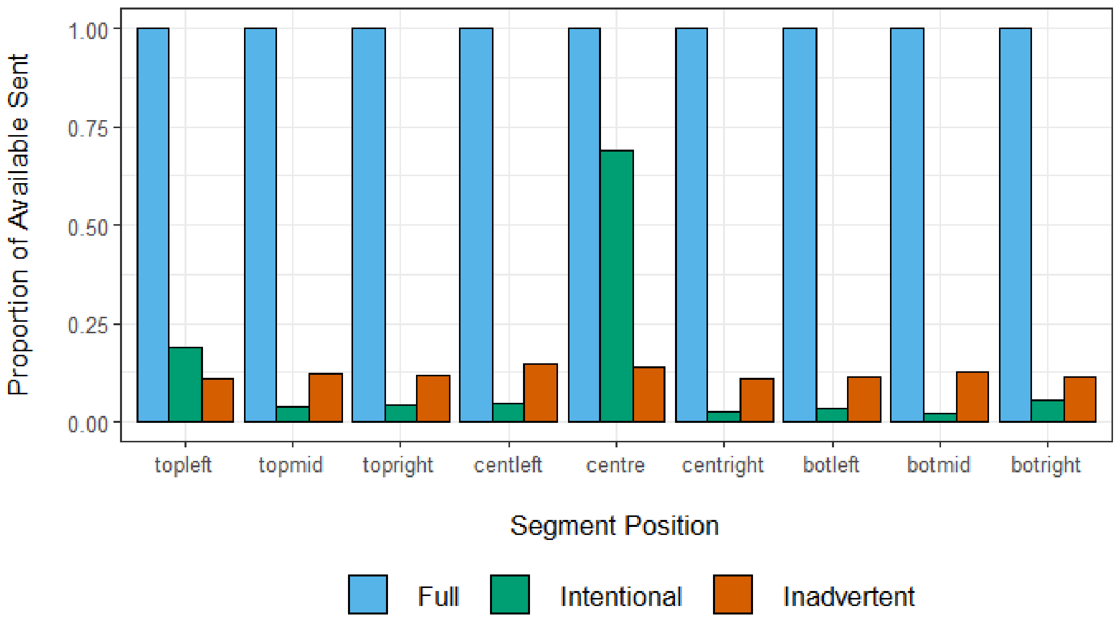
The positions sent within each target were analysed to determine whether information senders sent any particular target segment more frequently. The proportion of each target segment sent can be seen in Figure 21. The data for this analysis only included target segments sent from complete targets (out of 1,500 targets available across all chains, chain positions and grids in each conditions: Full condition, n = 1,085; Intentional condition, n = 805, Inadvertent condition n = 652).

To explore whether central tiles were transmitted more often relative to other target segment positions, we first compared the proportion of each target segment selection to a chance level selection. This analysis takes into account that there is only one central segment and eight edge segments in each target, and therefore this chance level was set to 0.11.

A Pearson's Chi-Square test was used to analyse any bias in the distribution of possible locations within targets (e.g., top left, centre, bottom right) sent in the Intentional and Inadvertent conditions. The Intentional condition was significantly different from a uniform distribution ($x^2(8)$ = 792.66, p < .001, n = 921), and the Inadvertent condition was not ($x^2(8)$ = 2.444, p = .964, n = 717).

Figure 21

The proportion of each target segment sent in each condition



Note. This graph incudes data from only completely uncovered targets. This graph therefore does not include any data from partially uncovered targets. X-axis values correspond to particular target segments. For example, '*topleft*' refers to the top left segment in a target. In this diagram, '*centre*' in shortened to '*cent*', and '*bottom*' is shortened to '*bot*'.

3.5 Discussion

The aim of this study was to investigate whether intentional knowledge sharing could support cumulative cultural evolution under circumstances of limited information transmission. Specifically, we aimed to investigate whether this would occur when the amount of information transmitted was so limited that it would otherwise severely constrain the retention of beneficial information (i.e., when it was available only in the form of inadvertent cues). We were therefore particularly interested in how our two limited information conditions (Intentional and Inadvertent) compared to a Full Information condition, in which participants received all of the information generated by their cultural parent's activity.

We found that task scores increased over learner generations, consistent with the accumulation of information over successive generations of learners. Overall task scores were significantly higher in the Full information condition, compared with the Inadvertent information condition, confirming that limiting the amount of information transmitted (but selecting this subset at random) restricted the potential for cultural offspring to benefit from the cumulative exploration activities of their cultural parents in the chain. However, in line with our key prediction, task scores were significantly higher in the Intentional condition compared with the Inadvertent condition. Furthermore, the difference between the Full and Intentional conditions was not significant. This result implies that – at least under certain circumstances – intentional knowledge sharing can fully compensate for the loss that would otherwise occur as a consequence of a tight bottleneck on transmission (e.g., due to limited exposure time).

The difference between the Intentional and Inadvertent conditions occurred because participants in the Intentional condition were apparently able to strategically select elements of their search activity for transmission, which would be particularly informative for the cultural offspring. In contrast, in the Inadvertent condition the information consisted of a random sample of the cultural parent's search activity. This suggested that participants in the Intentional condition may have been anticipating the needs of cultural offsprings, and that as a result they made significantly non-random choices about what to transmit in ways that did indeed benefit cultural offsprings.

Exploratory analysis of the information transmitted in the Intentional condition indicated that participants selected a disproportionate number of hits for transmission. Compared with information about miss locations, information about hit locations was strategically more useful to a cultural offspring aiming to maximise their score for a variety of reasons. Most obviously, using information about a hit (by selecting the same grid square) guaranteed that the cultural offspring could add a point to their score, whereas using information about a miss (by avoiding it and searching elsewhere) only very slightly increased the probability of scoring a point, relative to no information. In addition, the predictable clustering of target grid squares (the 27 hits within each grid were always arranged into three square targets of the same dimensions) meant that knowing the location of one target grid square narrowed down the potential locations of others that were part of the same target, making them much easier to find.

These biases in the information transmitted by participants in the Intentional condition (i.e., locations of hits rather than misses, and information distributed across all targets found) clearly allowed cultural offspring to perform extraordinarily well given the limited quantity of information they had about their cultural parent's search. Analysis of the performance of participants in the Intentional condition also showed that if they had been sent information by a participant who had found all three complete targets, they typically found all targets themselves before reaching their limit of failed search attempts (nine miss selections). This is quite remarkable given that participants only had information about the value of three of the 256 grid locations.

In contrast, in the Inadvertent condition, participants received information extracted randomly from the search of their cultural parent. This meant that even when all targets had been found, they might not all be represented in the information transmitted to the cultural offspring, either because one or more miss locations was sent, rather than a hit, or because the hit locations were not distributed across the three targets. As a result, in this condition, participants whose cultural parent had found all three complete targets performed less well than their counterparts in the Intentional condition, using up more of their search attempts, or failing to find all targets.

We can therefore state with some certainty that, in the Intentional condition, participants were selecting information for transmission in non-random ways. Furthermore, the resulting biases in the information provided did indeed benefit cultural offspring. However, the conclusion that this occurred as a consequence of the cultural parent's anticipation of the needs of the cultural offspring is at present only speculation. Further research would be required in order to investigate the motivations underlying the choices made by cultural parents in the Intentional condition. Such research could potentially add weight to the proposal that distinctively human intentional knowledge sharing can promote the accumulation of beneficial information.

Developmental research offers a potentially fruitful avenue for investigating the cognitive requirements of the strategic knowledge transmission behaviour we identified in the current study. Children's teaching behaviour is known to exhibit age-related changes consistent with an increasingly sophisticated understanding of others' minds (e.g., Ronfard & Corriveau (2016); Wood et al. (1995); Ziv et al. (2016)). Consequently, children's performance on this task could determine whether the biases in the intentionally transmitted information occurred as a result of such understanding. The ability to select appropriate information in the role of cultural parent might be age-dependent, and linked to the maturation of other cognitive abilities. This would add weight to the view that the highlighting of particularly informative behaviours could require cognitive capacities not available to non-humans, thus contributing to the explanation of why cumulative culture is rarely documented in other species. Alternatively, factors such as salience, or reinforcement history, could in principle produce effects similar to those identified in our experiment, without requiring any consideration of another's perspective. If even very young children make selections in the role of cultural parent that are comparable to those made by the adults, this would suggest that highlighting of details particularly beneficial to cultural offspring could itself occur inadvertently, as well as intentionally, potentially challenging the interpretation that such effects might be restricted to humans.

It should be noted that retention of information about target locations was far from perfect. This was the case across all conditions, including the Full Information condition, in spite of the extreme transparency of the task and the cues that were available within the experimental design. In the Full Information condition, participants were given information which revealed the value (hit or miss) of all of the grid squares searched by their cultural parent in the chain. Optimal use of this information virtually guaranteed improvement in score (or at least no deterioration) if all indicated hit locations were selected, and miss locations avoided. This was apparently trivial for most participants (as was our intention in planning the design). However, a minority of participants apparently either failed to understand the task rules and instructions, or adopted an extremely suboptimal game strategy (using up their full quota of moves searching non-cued locations prior to making use of the cues provided), or simply elected to ignore the instructed goal of maximising task score. These factors were outside of our control. However, although such behaviour violated some of our basic assumptions about how participants would respond to the task, we chose not to exclude any data on this basis, as this would have involved formulating post-hoc exclusion criteria. Given this context, it is particularly noteworthy that we identified the predicted differences between our conditions in spite of the resulting catastrophic collapses which wiped out the previously accumulated information. In both the Full and Intentional information conditions, chains were generally able to recover from such loss within just a few generations.

As a result of the catastrophic loss of information towards the end of chains, we included an additional exploratory analysis of performance up to where we considered the scores to 'plateau'. Identification of the plateau was based on real participant scores only, however, future research may benefit from including a simulated optimal performance model. This would provide an indication of whether participants are performing optimally, and also give a clearer insight into the point at which performance begins to deviate from this (i.e., the plateau point).

Overall, our findings suggest that intentional knowledge sharing can result in highly efficient transmission of beneficial information, such that cultural offspring do not necessarily need to be exposed to the full exploration history of their cultural parent in order to achieve the same, or better, performance. As a result, intentional knowledge sharing could be expected to significantly facilitate cumulative culture relative to learning from public information available in the form of inadvertent cues. Indeed, the efficient transmission of information is likely particularly critical to cumulative culture for the very reason that we expect cumulative culture to result in an increasingly large pool of discoveries worth transmitting, and/or increasingly hard-to-learn traits (e.g., see Mesoudi (2011)). The human propensity for intentional knowledge sharing may therefore play a key role in supporting cumulative culture, which is particularly beneficial once multi-generational transmission is already under way and has begun to produce traits that would be difficult for an individual to discover through trial and error.

Chapter 4: The development of intentional information sharing that promotes cumulative culture

4.1 Abstract

In the previous study, we found that intentional information transmission from cultural parents facilitated the accumulation of beneficial information over generations. Due to humans' distinct propensity for intentional information transmission, we concluded that this (relative to inadvertent information making up non-human transmissions) may be a key facilitator of distinctively human cumulative culture. In the previous chapter, adult's selective information transmission was speculatively attributed to a fully developed capacity for mental state understanding, such that they selected information for transmission that would be particularly beneficial to a naïve offspring. However, it is not completely clear from the results that adults' performance in the task was due to complex cognition, over simple biases and preferences. The current study therefore aimed to clarify that the effect found in adults was brought about by complex cognitive reasoning that is not accessible to non-humans and young children. If the same result that was found in adults is also found in young children, then it is unlikely that any complex reasoning is taking place. Specifically, the current study aimed to provide evidence that cumulative culture, supported by particularly beneficial information transmission, emerges on a similar developmental trajectory as advanced, and higher-order, mental state reasoning. We applied an experimental paradigm adapted from the one used in Chapter 3 with chains of children split by age: 5 to 6 years, 7 to 8 years and 9 to 10 years. We found that chains of 5- to 6-year-olds did not accumulate information in any condition; and 7- to 8-year-olds accumulated information only when all of the information from the cultural parent was available to the cultural offspring However, only 9- to 10-year-olds' performance resembled that of the adults, with this group exhibiting accumulation in the intentional transmission chains also. Children's cognitive abilities expand in a wide variety of ways over the age range we studied, and many of

these, not just capacities for mental state understanding, could have potentially underpinned the effects we found. The test for mental state understanding we had hoped to use as a predicting variable in our analyses returned results that suggested low validity, meaning that these scores could cast no further light on this issue. Nonetheless, the developmental trajectory we identified was consistent with the idea that the cognitive mechanisms responsible for the effect we found in the adult task were non-trivial and might be distinctive to humans. Furthermore, an ability to reason about others' minds remained amongst the most logically plausible explanations.

4.2 Introduction

In the previous chapter, we found that chains of adult participants transmitting subsets of intentional, compared to inadvertent, knowledge from cultural parents to cultural offspring demonstrated a significantly greater accumulation of knowledge over generations. Using an adapted version of the methodology applied in Chapter 3, the current study aims to assess whether the ability to intentionally select useful information for transmission is age-dependant and linked to the development of other cognitive abilities coming to fruition at the same time. Similarly to the previous study, we were particularly interested in whether any age-related changes in the ability to intentionally transmit beneficial information was reflected in the accumulation of information over multiple generations of transmission, compared with circumstances in which transmissions were made up of inadvertent information from the cultural parent's performance.

To effectively transmit knowledge (referred to on occasion in this chapter as '*teaching*'), there may often need to be some understanding from the cultural parent about where there are gaps in the cultural offspring's knowledge (Davis-Unger & Carlson, 2008). For this reason, in the conclusion of Chapter 3, we tentatively attributed the adults' ability to successfully transmit intentional knowledge and subsequently accumulate information over generations to distinctively human cognitive mechanisms, potentially including the fully developed understanding of others' mental states. However, if the same biases that were found in intentional adult transmissions (e.g., sending information about the centre grid square of fully revealed targets) are found in very young children, this would suggest that these biases are unlikely to be a consequence of sophisticated social cognitive mechanisms that only develop in later childhood. Of course, following a similar logic, if adult-like transmissions are apparent only in older children, then the age at which this develops would provide valuable insights into the level of social understanding required to effectively transmit knowledge in this task. In the current study we therefore attempted to capture a developmental trajectory for cumulative culture supported by intentional knowledge transmission, which could highlight whether this capacity can really be attributed to distinctively human cognitive mechanisms. Furthermore, we applied a test of mental state understanding suitable for children over the age of 4-years, to support the theory that the cognitive mechanism responsible is the fully realised understanding of other's mind states. Finding evidence of this would subsequently support the idea that intentional knowledge transmission may be dependent on capacities for reasoning about mental states, which are believed to be distinct in humans. This finding would also support theories present in the current literature, which often link the emerging capacity of mental state understanding with the ability to effectively teach others.

In the current literature, children's acquisition and understanding of teaching is consistently found to correlate with the developing knowledge of their own, and other's mental states, which are established at around 4 years of age (Chapter 1). Ashley & Tomasello (1998) and Wood et al. (1995) found that children rarely show any signs of peer-teaching (defined as circumstances where, during problem solving, a more knowledgeable child assists a naive child by adjusting their teaching strategy to suit the naïve child's needs) until around 3 years of age, where these abilities begin to emerge. After this age, children's repertoire of teaching behaviour increases, with experimental evidence citing children engaging in sibling/peer tutoring and cooperative learning from around 3 to 4 years of age (Ashley & Tomasello, 1998; Howe et al., 2012; Maynard, 2002; Wood et al., 1995).

However, it is not until the development of sophisticated mental state understanding at around 4 to 5 years of age, that children display a surge in teaching ability. At this stage, children begin to refine their information transmissions, using a larger rage of strategies, and responding more accurately to the learner's needs/errors (e.g., Davis-Unger & Carlson (2008); Gweon et al. (2014); Jeong & Frye (2018); Strauss et al. (2002); Ziv et al. (2008); Ziv & Frye (2004)). This is likely because, in order for a cultural parent to infer what a cultural offspring needs to know, some awareness of the offspring's knowledge state is required – more specifically, an ability to reason that gaps in the offspring's knowledge can potentially be addressed by a transfer of relevant information from the cultural parent.

Of course, to transfer the relevant information, the cultural parent must recognise that they have the knowledge in which to do so, and that what they transmit results in the cultural offspring gaining some knowledge that they did not previously have. Davis-Unger & Carlson (2008) found that children's responses to metacognitive questions indicated that their understanding of their own power as an informed cultural parent advanced with age, such that they were able to reason how they knew that the cultural offspring had learned a task as a result of the information they had transmitted as a cultural parent by around 4.5 years of age. Such reasoning indicates the ability to apply a capacity greatly linked to mental state understanding; sophisticated *level-two* perspective taking, i.e., the ability to reason about *how* knowledge or beliefs about a common object or concept may differ between agents. Level-two perspective taking emerges in elicited tasks at around the age of 3 to 4 years of age (Pillow & Flavell, 1986) and continues to mature until around 8 years of age (Salatas & Flavell, 1976).

Any perspective taking before 3 years of age tends to follow the model for *level-one* perspective taking, which only allows identification of common knowledge (i.e., what I can see, you can see) without any reasoning about differing knowledge or beliefs (Flavell et al., 1981). In Ashley & Tomasello (1998)'s task, children under 3 years of age did have some, albeit a limited, sense that their communicative partner had a different role, however, after 3 years of age, children began to develop a good grasp of their communicative partner's point of view, and what they could do as a cultural parent, to change it. This suggests that children over the age of 3 years were beginning to access level-two processes akin to adult-like mental state understanding and using this newfound capacity to support their teaching ability.

However, as aforementioned, level-two perspective taking is found to continue to develop long after the recognised developmental milestone at 4 years of age. This developmental pattern is very common for higher-order capacities related to mental state understanding, meaning that children continue to gain additionally complex belief reasoning throughout childhood (discussed in more depth in Chapter 1). This increase has been found to benefit the selective transmission of information to others.

One of the main questions in the current study relates to children's ability, as a cultural parent, to select information that is particularly useful for a cultural offspring. There are relatively few studies exploring how young cultural parents identify information worthy of transmission. To the best of our knowledge, there is only one such study: Ronfard et al. (2016) studied children aged 4 to 7 years old, and found that children from 4 years of age preferentially transmitted information that they themselves had acquired from a cultural parent, compared with information they acquired through individual exploration. This preference was only visible when the transmittable information is deemed '*more challenging to use'*, or '*difficult to acquire'* through individual exploration. In contrast, when difficulty for use and acquisition were equal for both information transmitted from a cultural parent and information that was self-discovered, children were relatively less likely to transmit a method they had acquired from the cultural parent. This means that children were specifically selecting information for transmission if it was deemed that they (and therefore also potentially the offspring) could not have acquired that knowledge through individual exploration.

Of course, the cost of both selecting information for transmission, and processing that information is high, and taking these costs into account is important when deciding what to transmit to a cultural offspring. For example, it is not beneficial to either party to teach more information than necessary, because this would require the cultural offspring to sift out what information is necessary to complete the desired task. Therefore, selective transmission of knowledge is required – particularly in circumstances where transmission is limited, as it is in the current task. This level of efficient teaching has been reported to increase between 3 and 7 years of age (Ronfard & Corriveau, 2016; Wood et al., 1995; Ziv et al., 2016). To transmit knowledge efficiently in this way, children must assess the cost of transmission relative to the benefit to the cultural offspring. Children above 5 years of age have been found recognise the learner's expected utility (Bridgers et al., 2016), for example, choosing to transmit difficult-to-learn methods for a return of the cultural offspring gaining a better reward. Children at this age will also transmit information at the expense of their own utility as long as the goal is being met (Gweon et al., 2014). This suggests that older children are able to make an accurate assessment of their own motivations and ability to transmit information.

Along with the cost of producing the information to be transmitted, the requirement to understand the cultural offspring's knowledge state means that information transmission on part of the cultural parent may carry quite a heavy cost. Judging by the adult participants' performance in Chapter 3, this may be an essential element in the current task, because cultural parents transmitted information that allowed the cultural offspring to make correct inferences about the location of adjacent rewards. It has been argued that the adaptivity of information transmission depends on evaluation of whether the cultural offspring could acquire the information through exploration. It is potentially wasteful to transmit something that the cultural offspring could discover through trial and error in a short time, however to the benefits are more likely to outweigh the costs when transmitting something that the cultural offspring would be unable to discover on their own (Csibra & Gergely, 2009; Fogarty et al., 2011).

Therefore, as in the adult task (Chapter 3), in cases where a cultural parent can only send a limited amount of information to an cultural offspring, selecting the information that is the most useful to the cultural offspring will be the greatest support in terms of cumulative culture. Doing this requires the cultural parent to have some knowledge of the cultural offspring's goals, current knowledge, and required knowledge. This thesis has already shown that adults can intentionally transmit information with these requirements in mind (Chapter 3), but the development of selective teaching to support cumulative culture has not been explored empirically.

The current study aimed to assess age-related changes in the ability to intentionally select information from one's own knowledge in a way that promotes an increase in acquired information over generations of transmission. We applied an experimental paradigm adapted from the one that was used in Chapter 3. The condition of intentional knowledge transmission was directly compared to a condition in which knowledge from a cultural parent was acquired incidentally from the parent's activity (reflecting the vast majority of social information use in non-humans, discussed in depth in Chapter 3).

As aforementioned, we were particularly interested in whether transmission chains of younger children were able to accumulate information in circumstances where intentional information sharing was permitted. If this were to be the case, this would call into question the claims made in Chapter 3, that accumulation over generations can be attributed to the cultural parent's sophisticated and distinctively human cognitive capacities. However, if chains of younger children are found not to benefit from intentional information sharing, this would add weight to the claims that cumulative culture may be distinct in humans as a result of similarly distinct cognitive prerequisites such as advanced mental state understanding. Based on the evidence supporting the relatively late development of mental state understanding, and its relationship to age related changes in teaching, we test the hypothesis that age-related changes will occur, such that younger children, compared to older children, will be unable to utilise intentional knowledge sharing to benefit accumulation over generations.

To further investigate the mechanisms responsible, we also used an established measure of mental state understanding, testing all children prior to their completion of the main grid search task. Evaluating advanced capacities for mental state understanding in older children is more challenging than answering the more straightforward question of whether a child has a concept of mental states *at all*. As a result there are very few suitable tests for this in the literature, and the validity of many of these is debatable (Quesque & Rossetti, 2020). We opted for the '*Yoni task'* (Shamay-Tsoory & Aharon-Peretz, 2007). The primary reason why we chose this measure over other established measures (such as the '*strange stories'* tasks (Happe et al. 1994) was for ease of administration, particularly because we could scale down the Yoni task to fit within the time we were permitted to have children outside of their classroom. Furthermore, as well as measuring both first- and second-order mental state understanding, it also measures reasoning about metacognitive belief and intention (cognitive aspects) and reasoning about emotional mental states (affective aspects).

Therefore, using this measure meant we could assess multiple factors in a relatively short time frame.

To the best of our knowledge, there are no studies in the current literature that have used the Yoni task as a measure of the development of mental state understanding specifically. However, its capacity to measure the multiple factors discussed above has been useful in assessing specific benefits of intervention strategies for children with impaired social cognition. For example, Kim et al. (2016) used 7- to 18-year-old children's scores in the Yoni task as a measure to assess how children with Autism Spectrum Disorder (ASD) compensate for an impaired mental state understanding. The Yoni task has also been used in 6- to 12-year-old children to assist in identifying suitable stimulants for increasing social cognition in children with ADHD, with children scoring higher in trials following intervention compared to scores pre-intervention (Maoz et al., 2014). These studies indicate that this measure is suitable for use with developmental populations.

While this task has not been used extensively with children, it has been used to measure mental state understanding capacity in adult populations, particularly comparing neurotypical adults and older adults that may have an impaired higher-order understanding of mental states (Fischer et al., 2017; Rossetto et al., 2018). Because children are said to develop a capacity for second-order mental state understanding by the age of 7 years (Perner & Wimmer, 1985), we saw no reason as to why this test would not show the emergence of higher-order mental state understanding if administered correctly. We hoped that this measure would produce scores that could potentially be used as a predictor in our analyses, and that any age-related changes in performance on this task might correspond to age effects identified in the grid search task.

For the main task, as in the previous chapter, participants took part in a search task which required them to locate hidden targets within landscapes segmented with a grid. Grid squares could be searched one at a time, and participants scored points by finding target segments using a limited lifespan of a finite number of unsuccessful search attempts. Information transmitted between participants included the location of grid squares searched by the predecessor, and also whether this was a hit (part of a target) or a miss (not part of a target).

To assess information transmission conditions across ages, we split participants into three age categories and concentrated chains within these categories: 5- to 6-year-old, 7- to 8-year-old and 9- to 10-year-old. These age categories were chosen in order to assess how increased ability in the task emerges following the development of mental state understanding in the first instance. Specifically, we were interested in whether the increasing development of mental state understanding and perspective taking influences success in the older age groups relative to the younger age groups.

The procedure for the main task follows a similar structure as the previous main task (Chapter 3). However, we made some parameter changes to account for testing large numbers of children of all ages. These edits were necessary because all children were tested in a school setting, and we endeavoured to reduce the amount of time spent out of class. Furthermore, we had some concerns about the concentration span of younger children and wished to limit the potential for fatigue effects. The main change we made in this respect was to reduce the number of grids given to each child to complete from ten to five to minimise the amount of time spent testing each child. To reduce the number of participants required we also reduced the length of transmission chains. Given that in the adult task we found that the scores achieved in chains plateaued earlier than we expected (around chain position six), we chose to reduce the size of chains from ten to five participants. We also made necessary edits to the instructions to make them more suitable for children to understand. The content of the instructions remained the same but some of the wording was made simpler, for example, *'previous participant'* was changed to '*previous space explorer'*. The script read to children can be seen in Appendix 7. To ensure children understood the instructions, an understanding check questionnaire was included. Finally, some aesthetic changes were made to the task; however, these had no effect on the task structure. Further details about the changes made to the task can be found in the procedure section.

In each age category, we ran 15 chains each composed of five participants, with five chains in each of three conditions: Full, Intentional and Inadvertent information. In the Full information condition, participants were shown all of the search feedback obtained by the previous participant. In the Intentional information condition, each participant was asked to select a subset of the hits and misses they had revealed during in the search stage (three grid squares) for transmission to the next participant. In the Inadvertent information condition, an equivalent subset (three grid squares) was randomly selected from the feedback revealed during each participant's search, and this formed the information transmitted to the next participant in the chain (see Methods for full details).

We had no particular expectations about age effects in this task, other than that older children's performance would be the most similar to the results found in the adult task. While we were most interested in the difference between results in the Intentional and Inadvertent conditions, it was also of some interest to assess whether there was a particular age group that could accumulate information in the Full condition. This would highlight a potential for benefitting from information about the previous participant's performance, however in the absence of an advantage for intentional over inadvertent information transmission in the same age group, this would support the idea that the effect found in the adult task was due to the relative value of the information transmitted in the intentional condition.

4.3 Methods

4.3.1 Participants

Two hundred and thirty three child participants aged between 5 years, 0 months and 10 years, 11 months (Median = 7 years, 10 months, 116 females) were recruited in total, with data from six of these participants being excluded from the study due to experimenter error, technical error, or consent form errors (i.e., the parent had listed the child as being older or younger than their date of birth suggested, meaning that they did not belong in the category to which they had been assigned). The final sample consisted of 225 child participants aged between 5 years, 0 months and 10 years, 11 months (Median = 7 years 11 months, 113 females, for breakdown of age categories, see Table 3). Participants were recruited from six different primary schools in the north of Scotland, with the minimum number of children collected from one school being 13, and the maximum number being 46. Due to a technical error which was only discovered after data collection had been completed, the data from two of the participants from the 5- to 6-year age category was not fully recorded by the computer. However, for both of these participants, the information about their final scores and search attempts remaining had been recorded and therefore they were included in the primary analysis. However, because the data for which specific grid squares, they sent to the following participant were not recorded, this information was excluded from the relevant subsections of the analysis.

Table 3

Descriptive statistics for each age category

Category	Minimum	Maximum	Mean	Median
5 to 6 years	5 years, 0 months	6 years, 11 months	6 years, 1 month	6 years, 1 month
7 to 8 years	7 years, 0 months	8 years, 11 months	7 years, 11 months	8 years, 0 months
9 to 10 years	9 years, 0 months	10 years, 11 months	9 years, 11 months	10 years, months

75 participants were assigned to each of three conditions (Intentional, Inadvertent and Full information), with 25 in each age category (5 to 6 years, 7 to 8 years, and 9 to 10 years). Five participants were assigned to each chain, hence there were five chains of each age category in each condition. Participants were assigned to a condition randomly, and chain position was constrained by the current state of completion of the chain to which they were recruited. Participants whose data was excluded were

replaced in that chain by a new participant, prior to further recruitment to that particular chain.

The study was approved by the University of Stirling General University Ethics Panel (approval reference number: GUEP 557). Formal consent to conduct the study was obtained from The Highland Council education manager, and all head teachers gave consent for the experimenter to enter the school at least two weeks in advance of testing. Signed parental consent was obtained prior to testing which confirmed that children were permitted to take part in the experiment, and that this could be done during class time. Participants and their teachers also gave verbal consent before the participant was removed from their classroom.

4.3.2 Mental state understanding test

Before completing the experimental task, participants were given a test of mental state understanding adapted from Shamay-Tsoory & Aharon-Peretz (2007)'s '*Yoni task'*. This task was run using PsychoPy v1.84.2 (Peirce et al., 2019), on either a Microsoft Surface tablet, or a Lenovo Yoga 520 touchscreen laptop in tablet mode, both using Windows 10.

All participants completed the mental state understanding test, however data from one participant in the 5- to 6-year-old age category and one participant from the 7- to 8-year-old age category failed to record due to a technical error.

Due to the adaptions we made to the task, and based on feedback received during informal piloting, we re-named the main protagonist in the task '*Yanny'*. Therefore this task will be referred to as the '*Yanny task'* from this point forward. As in the original Yoni task, the Yanny task included separate trials testing for first- and second-order mental state understanding. All trials showed a yellow character (named '*Yanny'*) in the middle of a screen, a different image in each corner of the screen, and an incomplete sentence about Yanny at the top of the screen (as shown in Figures 22a and 22b). We reduced the number of trials in the original task so we could fit it within the timeframe given to test each child, however, we ensured all of the essential elements of the original task were retained, such as at least one of each trial type. All of the stimuli images were re-created because the original task images contained some ambiguity about where some characters were looking, which we thought likely to affect children's ability to successfully select target images. Otherwise, the layout of the task was the same.

Children were asked to complete sentences that are assumed to require inference about a character's cognitive, physical or affective state. In the First-order trials the child had to make a judgement about Yanny's state in relation to particular objects. However in the Second-order trials the child had to make a judgement about Yanny's state in relation to objects that could only be identified with reference to another character's state. All of the sentences children were asked to complete can be seen in the stimuli shown in Appendix 5 and 6.

Figure 22a

An example of a trial assessing First-order mental state understanding.

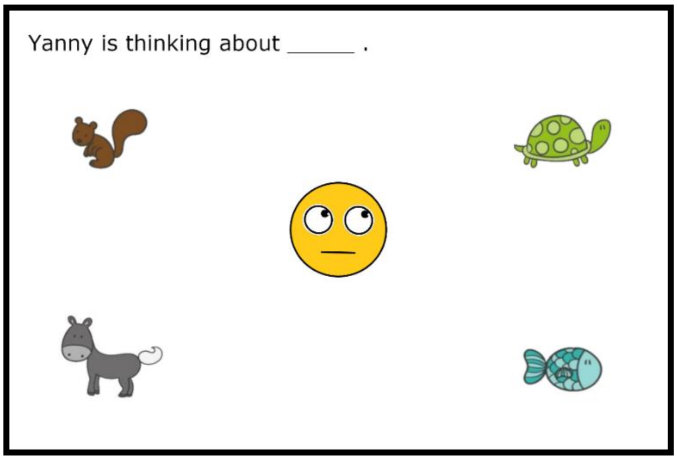
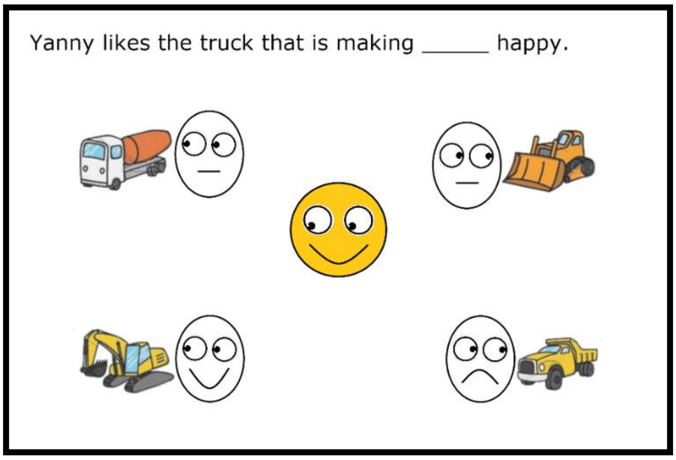


Figure 22b

An example of a trial assessing Second-order mental state understanding.



Note. Children were read the sentence at the top of the screen and were asked to select which of the four other illustrations could correctly complete the sentence.

4.3.3 Procedure: Mental state understanding test

Children were read brief instructions (Appendix 5 and 6) which introduced the Yanny character and explained that they were to select the object or person that Yanny was having a thought or a feeling about. Following this, children were shown 24 images which were headed with an incomplete sentence that could be completed by selecting one of four stimuli situated in the corners of the screen.

An experimenter read aloud the sentence shown at the top of the screen and asked the child to select which of the four objects would correctly complete the sentence. Once the child selected an object, the next trial was shown. Participants completed 12 First-order trials in the first stage, all of which were assumed to require first-order mental state understanding to complete (e.g., "*Yanny likes*____"). Following this, children completed 22 Second-order trials in the second stage, all of which were assumed to require second-order mental state understanding to complete (e.g., "*Yanny likes the animal that* ____ *does not like*"). All children completed the blocks and trials in the same order. The reason for running the stages in the same order was based on a pragmatic decision to test all children using the easiest condition first (the First-order block). In the lead up to running this experiment it was decided that running the trials in the same order would be the easiest option, however in hindsight randomizing the order of the trials would have avoided any potential confounds caused by practice effects.

Nonetheless, the potential for practice effects was relatively limited. The experimenter did not react to the child's selection, and there were no built-in indicators of success. Therefore, children were given no feedback regarding whether or not they had made the '*correct'* choice. This was in order to reduce reinforcement effects.

4.3.4 The experimental task

The task completed by the participants was a landscape searching challenge loosely based on the game '*Battleships'*. This task was run using PsychoPy 1.84.2 (Peirce et al., 2019), on either a Microsoft Surface tablet, or a Lenovo Yoga 520 touchscreen laptop in tablet mode, both running Windows 10.

Participants were presented with a series of five 16x16 grids, in each of which three 3x3 targets (ships) could be found. Targets were randomly placed, and consisted of nine grid points (See Figure 23). The reward landscape was randomised for each chain, such that within chains, participants completed the same five grids. A new random set of grids was generated for each chain, meaning that there were 45 new sets of grids generated in total. Participants could search the grid for targets by touching any of the grid squares. When part of a target was found, these were scored as '*hits'*, and selection of any of the non-target grid squares in the 16x16 grid were scored as '*misses'*. The participant's primary goal was to maximise their score by finding as many hits as possible, across all five grids.

Participants used up one of their limited search attempts for every grid square they selected which did not contain part of a target. In contrast, when participants selected a target grid square, the number of search attempts did not reduce, and they were instead awarded one point. This scoring system was intended to capture the relative payoffs typically associated with real-world subsistence activity, with fruitless search always costly, and success always profitable. A maximum score of 27 points was available in each grid (nine for every full target), and participants were allotted nine search attempts in which to find them. This meant that the score for each grid varied depending on how many hits and misses were found. For example, if all 27 target points were found without any error, the score for that grid would be 27, with nine search attempts remaining and exactly 27 grid squares revealed. If no targets were found, the score for that grid would be zero, with all search attempts used and only nine (miss) grid squares revealed.

The first participant in each chain received no information about the contents of any of the grid squares, other than that generated by their own search. All other participants in the chain received some information about their immediate predecessor's performance for each grid.

The same pre-specified reward landscape was kept consistent for each grid completed by a particular chain, such that the results of the predecessor's search held true for the successor. Depending on the condition, participants saw some or all of the outcome of their predecessor's search (including whether each revealed grid point was a hit or a miss). The nature of the information varied by condition as detailed below:

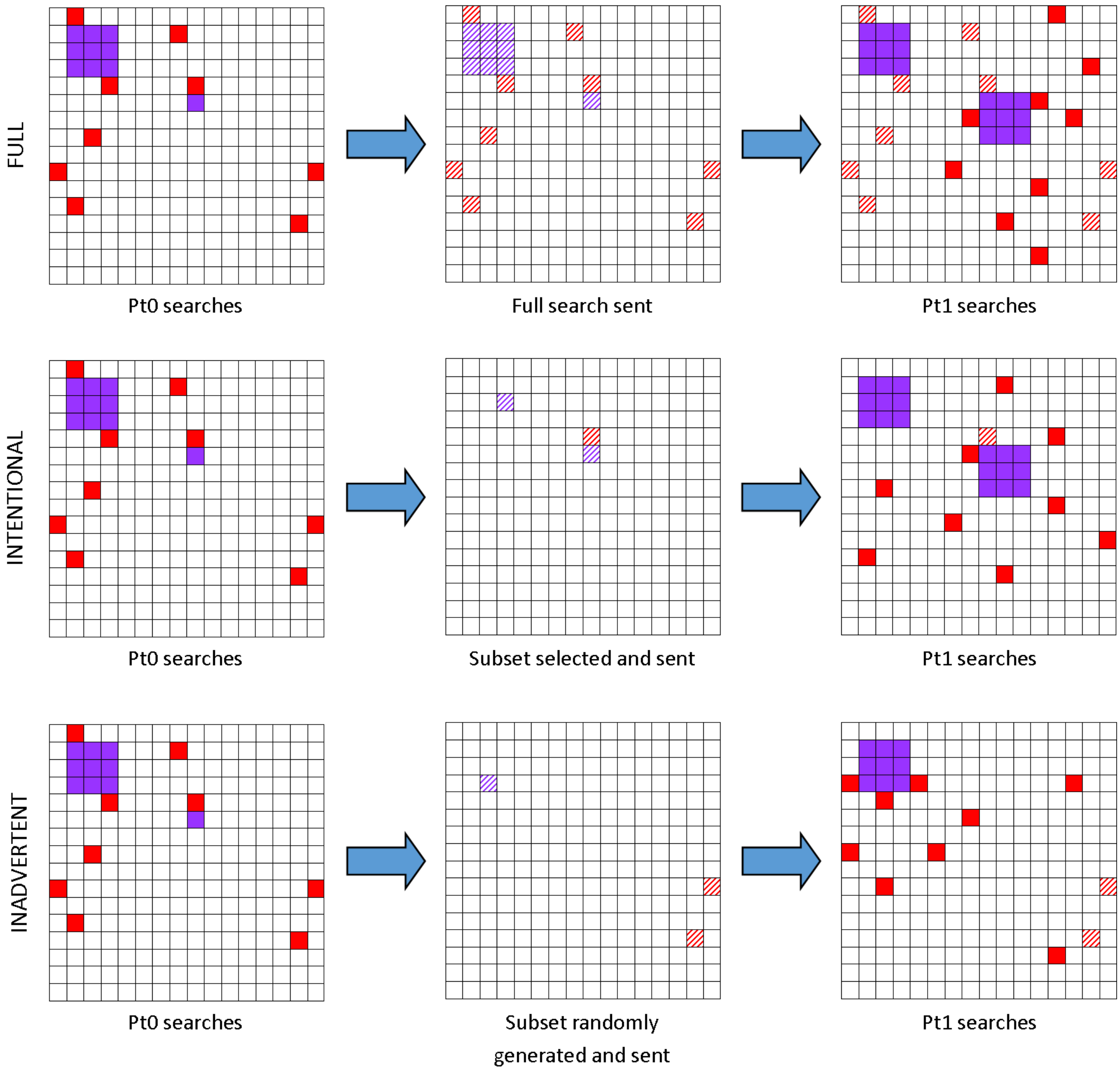
Full Information (control): With the exception of the first in each chain, each participant was shown all of the search feedback obtained by the previous participant. Thus, participants received information about the value of between nine and 36 grid squares.

Intentional Information: At the end of their search, each participant in this condition was asked to select a subset of the hits and misses they had revealed during the search stage (three grid squares). Thus, the information that each participant received comprised the value of these three squares, as selected by their predecessor in the chain.

Inadvertent Information: A randomly selected subset (three grid squares) of the grid squares revealed during each participant's search were passed on to the next participant in the chain. Thus, as in the Intentional condition, participants in this condition received information about the value of three grid squares. This condition was therefore intended to be analogous to using '*public information*' (e.g., Danchin et al. (2004)), available as a by-product of another's activity, as long as the observer happens to be in the vicinity of the information producer.

Figure 23

Demonstration of information transmission between two generations.



Note. Filled purple grid squares indicate a hit, filled red grid squares indicate a miss. Striped grid squares in the rightmost grids indicate information that was transmitted from the cultural parent but not subsequently selected by the cultural offspring. Blue arrows represent the direction of transmission.

4.3.5 Procedure: The experimental task

Following the consent procedure, participants were pseudo-randomly assigned a condition and chain position according to the constraints detailed in the Participants section.

To make the task more appealing to children, we changed the task stimuli from '*battleships*' to '*spaceships*'. This aesthetic change only included switching the colour of hits from green to purple, the hit noise from an '*explosion*' to an '*alien blip*', and the background from plain grey to a colourful space-themed scene.

Participants were provided with thorough instructions on screen before beginning the experiment. These instructions were adapted from the adult task instructions to be easier for children to understand. These instructions included picture examples of an unselected grid square, a selected hit, a selected miss and a full 3x3 target. Participants were also given instructions about the constraints on their search activity for any given grid, i.e., that the program would advance to a new grid when either all nine search attempts had been used up, or when all 27 scoring tiles (i.e., three complete targets) had been selected (see description of experimental task in section 4.3.4). Following the instructions, participants were presented with an understanding check quiz (Appendix 4) that ensured they could identify a 3x3 target, recall how many full targets were hidden, and, with the exception of the first participant in the chain, remember which transmissions from a previous participant indicated a hit and a miss. Children were given the quiz as soon as the instruction screens ended, and were encouraged to discuss the instructions until they had correctly answered all of the questions. The quiz was repeated at the end of the task to ensure they were still aware of the instructions. Children could refer to the quiz for reminders of the shape of the targets throughout the task.

If a participant was in generation 1 (i.e., the first participant in a chain of five), then they received no information about the location of hits or misses. Participants in all other chain positions received some information about the grid, as detailed in the description of the experimental task. In all conditions, participants were told that they would receive *some information* from a previous participant's performance, and that they could use this information however they wished. Details about the source of the information was not explicitly disclosed to participants.

In the event that participants stalled their search due to distraction, the experimenter gave them a prompt (such as "*find the alien spaceship*"). The full list of prompts used can be seen in Appendix 3. After participants had completed their own search, in the Intentional condition, they were also required to make decisions about the information to be transmitted to the next participant in the chain. Participants in this condition were therefore made aware that the information was to be sent to a future participant who would be presented with the same task they had just completed. Following each grid search, they were asked to select three of the grid squares from the selections that they had made for any given grid.

After all five grids had been searched, participants were shown their final score (total hits selected across all five grids). Participants were given the opportunity to choose a reward sticker for participating.

4.4 Results

All generalised linear mixed models were performed using R (R Core Team, 2020) and lme4 (Bates et al., 2014). Graphs were created using *ggplot 2* performed on R Studio (RStudio Team, 2020), supported by *R version 4.0.2*.

Non-convergent and singular fit models were addressed by removing random variance, however models with the maximum random effects structure (Barr et al., 2013) were considered first. Post-hoc comparisons were done using the emmeans package (Lenth et al., 2019). Tests were two-tailed, and p-values of $< .05$ were taken as statistically significant. Emmeans using the Tukey adjustment was used for multiple comparisons, and emtrends was used to assess interactions between slopes when significant interactions were found in the model.

4.4.1 Mental state understanding test

The proportion of correct responses recorded in each trial can be seen in (Figure 24).

Generalized linear models were conducted to assess possible practise effects in the First- and Second-order stages. Each stage was conducted as a separate model. Children's response to trials was included as a binary dependant variable.

In both models, a significant main effect of trial number was found (First-order: (b = 0.015, SE = 0.002, $t(2674)$ = 7.035, $p < .001$); Second-order: (b = 0.004, SE = 0.0006, $t(4904)$ = 5.632, $p < .001$)). This suggests that practice effects were present in both conditions.

A general linear mixed effects model was conducted using stage and age category as fixed effects, including stimulus type (physical, cognitive, affective), and trial number as random effects. A significant main effect of stage was found (b = 0.423, SE = 0.119, z = 3.544, p = < .001).

Emmeans was used to explore this main effect further, firstly comparing the differences between stages within each age category. Scores in the First-order condition were significantly lower than in the Second-order condition in the 5- to 6-year age category (b = -0.423, SE = 0.119, z = -3.544, $p < .001$), in the 7- to 8-year age category (b = -0.890, SE = 0.134, z = -6.652, $p < .001$), and in the 9- to 10-year age category (b = -0.318, SE = 0.137, z = -2.328, p = .020).

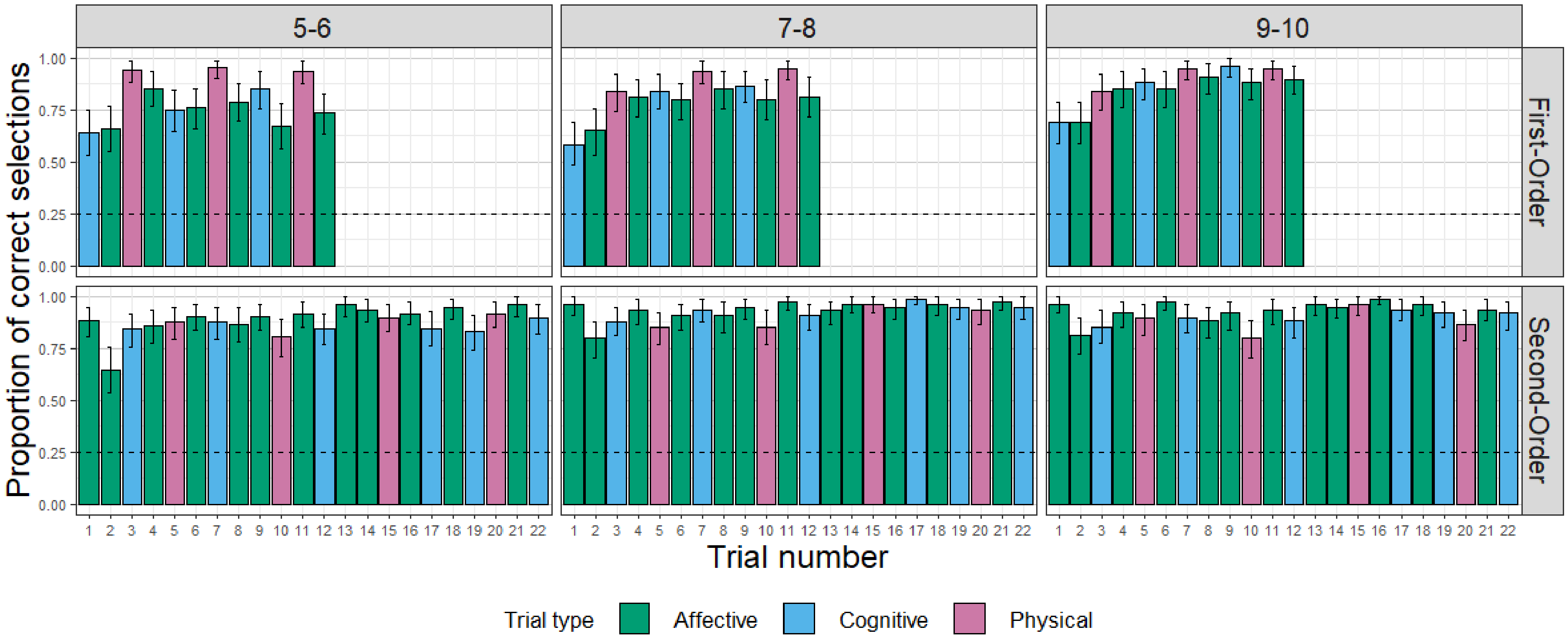
Following this, emmeans was used to compare differences between age categories within stage. In the First-order condition, scores were significantly higher in the 9- to 10-year age category compared with the 5- to 6-year age category (b = 0.518, SE = 0.129, z = 4.006, $p < .001$), and in the 9- to 10-year age category compared with the 7- to 8-year age category (b = 0.400, SE = 0.131, z = 3.051, p = .006). Scores in the 5- to 6-year age category were not significantly different from scores in the 7- to 8-year age category (b = -0.118, SE = 0.121, z = -0.972, p = .594). In the Second-order condition, scores were significantly higher in the 7- to 8-year age category compared with the 5- to 6-year age category (b = 0.585, SE = 0.122, z = 4.796, $p < .001$), and in the 9- to 10-year age category compared with the 5- to 6-year age category (b = 0.413, SE = 0.117, z

= 3.453, p = .001). Scores were not significantly different in the 7- to 8-year age category compared with the 9- to 10-year age category (b = 0.172, SE = 0.130, z = 1.323, p = .383).

In this analysis, neither the expected age effects, or stage effects were found, and it is likely that this was largely as a result of performances close to ceiling in all groups for both stages. This gave a good indication that the test was not providing a sensitive measure, and that it was failing to pick up individual differences in advanced mental state understanding, rendering it of little value to use as a predictor in our analyses of the main task. For this reason, the results of this task were not used to bolster claims about the facilitatory effect of mental state understanding in any other aspects of the task.

Figure 24

Proportion of correct responses recorded in the First-order and Second-order mental state understanding test trials.



Note. Split by age category and stage. Error areas represent +/- 1 standard error

4.4.2 Mean scores and search attempts remaining over generations of learners.

First, we examined the scores at chain position 1 using condition as the independent variable, and chain included as a random variable. The Intentional condition was used as the intercept. We conducted separate analyses for scores and search attempts remaining, and within each we conducted models comparing age within condition, and condition within age (12 separate models). Scores and search attempts remaining at chain position 1 were all non-significantly different, apart from scores in the Full condition between the 5- to 6-year age group scored significantly lower than the 9- to 10-year age group (b = -15.0, SE = 5, $t(12)$ = -2.998, p = .028). However, if the appropriate Bonferroni corrections were run, taking into account that this would include 12 tests, this would unquestionably be returned as non-significant.

4.4.2a Comparing condition within age category

To assess the differences between the conditions in each age category, we split the data by age category. We conducted three models with '*score'* as a dependant variable, and three models with '*search attempts remaining'* (both numeric) as the dependant variable.

Mean scores

A separate model was run for each age category. Therefore, in each model, only the data from the specified age category was used in order to compare the differences between conditions within that age category alone. In these models, we included condition and chain position (centred) as independent variables, chain is included as a random variable where possible. The Intentional condition was used as the intercept (Figure 25).

5- to 6-year-old chains:

For the dependent variable of score, a significant main effect was found for chain position (centred) in the 5- to 6-year-old chains (*b* = 34.700, *SE* = 13.964, *t*(69) = 2.485, *p* = .015).

In the 5- to 6-year-old chains, scores were not significantly different in the Full condition when compared to the Intentional condition (*b* = 3.56, *SE* = 5.59, *t*(69) = 0.637, *p* = .800). Scores in the Full condition were not significantly different when compared to the Inadvertent condition (*b* = 7.36, *SE* = 5.59, *t*(69) = 1.318, *p* = .390). There was also no significant difference between scores in the Intentional condition compared with the Inadvertent condition (*b* = 3.80, *SE* = 5.59, *t*(69) = 0.680, *p* = .776).

In the 5- to 6-year-old chains, no significant interaction of score increase over chain positions was found between the Full and Intentional conditions (*b* = 14.8, *SE* = 19.7, *t*(69) = 0.749, *p* = .735) or between the Full and Inadvertent conditions (*b* = 25.0, *SE* = 19.7, *t*(69) = 1.266, *p* = .419). No significant interaction of score increase over chain positions was found between the Intentional and Inadvertent conditions (*b* = 10.2, *SE* = 19.7, *t*(69) = 0.517, *p* = .864).

7- to 8-year-old chains:

For the dependent variable of score, a marginally non-significant main effect was found for chain position (centred) in the 7- to 8-year-old chains (*b* = 29.600, *SE* = 14.914, *t*(69) = 1.985, *p* = .0511).

In the 7- to 8-year-old chains, participants' scores in the Full condition were significantly higher than in the Intentional condition (*b* = 16.72, *SE* = 5.97, *t*(69) = 2.803, p = .018). There was a marginally non-significant difference between scores in the Full compared to Inadvertent condition (*b* = 13.20, *SE* = 5.97, *t*(69) = 2.213, *p* = .076), and no significant difference between the Intentional and Inadvertent conditions (*b* = 3.52, *SE* = 5.97, *t*(69) = 0.590, *p* = .826).

In the 7- to 8-year-old chains, a significant interaction of score increase over chain positions was found between the Full and Intentional condition (*b* = 79.1, *SE* = 21.1,

$t(69) = 3.750$, $p = .001$) and a significant interaction was found between the Full and Inadvertent conditions ($b = 85.7$, $SE = 21.1$, $t(69) = 4.063$, $p < .001$). No significant interaction of score increase over chain positions was found between the Intentional and Inadvertent conditions ($b = -6.6$, $SE = 21.1$, $t(69) = -0.313$, $p = .948$).

9- to 10-year-old chains:

For the dependent variable of score, a significant main effect was found for chain position (centred) in the 9- to 10-year-old chains (chain included as a random variable) ($b = 61.900$, $SE = 9.457$, $t(74) = 6.545$, $p < .001$).

In the 9- to 10-year-old chains, participants' scores were higher in the Full condition compared with the Inadvertent condition ($b = 22.36$, $SE = 3.78$, $t(65) = 5.911$, $p < .001$). Comparable to the pattern identified in adults, scores in the Intentional condition were significantly higher than those in the Inadvertent condition ($b = 23.92$, $SE = 3.78$, $t(65) = 6.323$, $p < .001$), and scores in the Full condition were not significantly different than those in the Intentional condition ($b = -1.56$, $SE = 3.78$, $t(65) = -0.412$, $p = .911$).

In the 9- to 10-year-old chains, a significant interaction of score increase over chain positions was found between the Full and Intentional conditions ($b = 45.1$, $SE = 13.4$, $t(65) = 3.372$, $p = .004$) and between the Full and Inadvertent conditions ($b = 51.6$, $SE = 13.4$, $t(65) = 3.858$, $p < .001$). No significant interaction of score increase over chain positions was found between the Intentional and Inadvertent conditions ($b = 6.5$, $SE = 13.4$, $t(65) = 0.486$, $p = .878$).

Mean search attempts remaining

A separate model was run for each age category. Therefore, in each model, only the data from the specified age category was used in order to compare the differences between conditions within that age category alone. In these models, we included condition and chain position (centred) as independent variables, chain is included as a random variable where possible. The Intentional condition was used as the intercept (Figure 26).

5- to 6-year-old chains:

There was no significant main effect found for centred chain position in the 5- to 6-year-old age category (chain included as a random variable) (b = 1.80, SE = 1.248, t(74) = 1.441, p = .154).

In the 5- to 6-year-old chains, the sum of search attempts remaining was not significantly different between the Full condition and the Intentional condition (b = 0.48, SE = 0.5, t(65) = 0.961, p = .604), or between the Full condition and Inadvertent condition (b = 0.88, SE = 0.5, t(65) = 1.762, p = .191). The sum of search attempts remaining was also not significant between the Intentional condition and the Inadvertent condition (b = 0.40, SE = 0.5, t(65) = 0.801, p = .704).

In the 5- to 6-year-old category, no significant interaction of increase in search attempts remaining over chain positions was found between the Full and Intentional conditions (b = 1.9, SE = 1.77, t(65) = 1.076, p = .532), or between the Intentional and Inadvertent conditions (b = 2.0, SE = 1.77, t(65) = 1.132, p = .497). A non-significant difference was also found between the Full and Inadvertent conditions (b = 3.9, SE = 1.77, t(65) = 2.208, p = .077).

7- to 8-year-old chains:

There was no significant main effect found for centred chain position in the 7- to 8-year-old age category (chain included as a random variable) (b = 0.50, SE = 2.172, t(74) = 0.230, p = .818). However, in the 7- to 8-year-old chains, the sum of the search attempts remaining was significantly higher in the Full condition when compared to the Intentional condition (b = 9.28, SE = 0.9, t(65) = 10.681, p < .001), and significantly higher in the Full condition when compared to the Inadvertent condition (b = 9.12, SE = 0.9, t(65) = 10.497, p < .001). There was no significant difference between search attempts remaining in the Intentional condition compared with the Inadvertent condition (b = -0.16, SE = 0.9, t(65) = -0.184, p = .982).

In the 7- to 8-year-old chains, a significant interaction of increase in search attempts remaining over chain positions was found between the Full and Intentional condition (b = 25.9, SE = 3.07, t(65) = 8.432, p < .001) and between increase in search attempts remaining in the Full and Inadvertent conditions (*b* = 27.6, *SE* = 3.07, *t*(65) = 8.985, *p* < .001). No significant interaction of the increase of search attempts remaining over chain positions was found between the Intentional and Inadvertent conditions (*b* = 1.7, *SE* = 3.07, *t*(65) = 0.553, *p* = .845).

9- to 10-year-old chains:

There was a significant main effect found for centred chain position the 9- to 10-year-old age category (*b* = 13.90, *SE* = 2.504, *t*(74) = 5.551, *p* < .001). In the 9- to 10-year-old chains, the sum of search attempts remaining was significantly higher in the Full condition when compared to the Inadvertent condition (*b* = 7.12, *SE* = 1, *t*(65) = 7.108, *p* < .001), and significantly higher in the Intentional condition when compared to the Inadvertent condition (*b* = 4.80, *SE* = 1, *t*(65) = 4.792, *p* < .001). There was a marginally non-significant difference between search attempts remaining in the Full condition compared with the Intentional condition (*b* = 2.32, *SE* = 1, *t*(65) = 2.316, *p* = .060).

In the 9- to 10-year-old chains, a significant interaction of increase in search attempts remaining over chain positions was found between the Full and Intentional condition (*b* = 11.5, *SE* = 3.54, *t*(65) = 3.247, *p* = .005), between the Full and Inadvertent conditions (*b* = 25.2, *SE* = 3.54, *t*(65) = 7.116, *p* < .001), and between the Intentional and Inadvertent conditions (*b* = 13.7, *SE* = 3.54, *t*(65) = 3.869, *p* < .001).

4.4.2b Comparing age category within condition

To assess the differences between the age categories in each condition, we split the data by condition. We conducted three models with '*score'* as a dependant variable, and three models with '*search attempts remaining'* as the dependant variable.

Mean scores

A separate model was run for each condition. Therefore, in each model, only the data from the specified condition was used in order to compare the differences between age categories within that condition alone. In these models, we included age category and chain position (centred) as independent variables, chain is included as a random variable where possible. The 5- to 6-year-old age category was used as the intercept (Figure 25).

Full Condition:

A significant main effect was found for centred chain position in the Full condition (chain included as a random variable) (b = 49.5, SE = 10.46, $t(74)$ = 4.732, $p < .001$). In the Full condition, the sum score was significantly lower for 5- to 6-year-olds when compared to 7- to 8-year-olds (b = 29.12, SE = 6.82, $t(12)$ = 4.272, p = .003), and significantly lower for 5- to 6-year-olds when compared to 9- to 10-year-olds (b = -36.60, SE = 6.82, $t(12)$ = -5.370, $p < .001$). There was no significant difference between scores for 7- to 8-year-olds compared with 9- to 10-year-olds (b = -7.48, SE = 6.82, $t(12)$ = -1.097, p = .534).

In the Full condition, a significant interaction of score increase over chain positions was found between the 5- to 6-year-olds and 7- to 8-year-olds (b = -59.2, SE = 14.8, $t(57)$ = -4.002, $p < .001$) and between the 5- to 6-year-olds and 9- to 10-year-olds (b = -57.5, SE = 14.8, $t(57)$ = -3.887, $p < .001$). No significant interaction of score increase over chain positions was found between 7- to 8-year-olds and 9- to 10-year-olds (b = 1.7, SE = 14.8, $t(57)$ = .115, p = .993).

Intentional Condition:

A significant main effect was found for centred chain position in the Intentional condition (chain included as a random variable) (b = 34.7, SE = 13.752, $t(74)$ = 2.523, p = .014. In the Intentional condition, there was a non-significant difference between scores for 5- to 6-year-olds compared with 7- to 8-year-olds (b = -16.0, SE = 6.65, $t(12)$ = -2.399, p = .080). The sum score was significantly lower for 5- to 6-year-olds when compared to the 9- to 10-year-olds (b = -41.7, SE = 6.65, $t(12)$ = -6.272, $p < .001$), and significantly lower for 7- to 8-year-olds when compared to 9- to 10-year-olds (b = -25.8, SE = 6.65, t(12) = -3.872, p = .006).

In the Intentional condition, no significant interaction of score increase over chain positions was found between the 5- to 6-year-olds and the 7- to 8-year-olds (b = 5.1, SE = 19.4, t(57) = 0.262, p = .963), between the 5- to 6-year-olds and 9- to 10-year-olds (b = -27.2, SE = 19.4, t(57) = -1.399, p = .348). or between the 7- to 8-year-olds and 9- to 10-year-olds (b = -32.3, SE = 19.4, t(57) = -1.661, p = .229).

Inadvertent Condition:

A non-significant main effect was found for centred chain position in the Inadvertent condition (chain included as a random variable) (b = 24.5, SE = 12.914, t(74) = 1.897, p = .062). In the Inadvertent condition, the sum score was significantly lower for 5- to 6-year-olds when compared to 7- to 8-year-olds (b = -23.28, SE = 5.46, t(12) = -4.261, p = .003), and significantly lower for 5- to 6-year-olds when compared to 9- to 10-year-olds (b = -21.60, SE = 5.46, t(12) = -3.954, p = .005). There was no significant difference between scores for 7- to 8-year-olds compared with 9- to 10-year-olds (b = 1.68, SE = 5.46, t(12) = 0.308, p = .949).

In the Inadvertent condition, no significant interaction of score increase over chain positions was found between 5- to 6-year-olds and the 7- to 8-year-olds (b = 1.5, SE = 18.3, t(57) = 0.082, p = .996), between the 5- to 6-year-olds and 9- to 10-year-olds (b = -30.9, SE = 18.3, t(57) = -1.692, p = .217), or between the 7- to 8-year-olds and 9- to 10-year-olds (b = -32.4, SE = 18.3, t(57) = -1.774, p = .189).

Mean search attempts remaining

A separate model was run for each condition. Therefore, in each model, only the data from the specified condition was used in order to compare the differences between age categories within that condition alone. In these models, we included age category and chain position (centred) as independent variables, chain is included as a random variable where possible. The 5- to 6-year-old age category was used as the intercept (Figure 26).

Full Condition:

No significant main effect of centred chain position was found in the Full condition (chain included as a random variable) (b = 3.7, SE = 2.631, $t(74)$ = 1.406, p = .164).

In the Full condition, the sum of search attempts remaining was significantly lower for the 5- to 6-year-olds when compared to 7- to 8-year-olds (b = -8.80, SE = 1.57, $t(12)$ = -5.613, p < .001), and significantly lower for the 5- to 6-year-olds when compared to 9- to 10-year-olds (b = -6.24, SE = 1.57, $t(12)$ = -3.980, p = .005). There was no significant difference between search attempts remaining for the 7- to 8-year-olds compared with the 9- to 10-year-olds (b = 2.56, SE = 1.57, $t(12)$ = 1.633, p = .270).

In the Full condition, a significant interaction of increase in search attempts remaining over chain positions was found between the 5- to 6-year-olds and the 7- to 8-year-olds (b = -22.7, SE = 3.72, $t(57)$ = -6.101, p < .001) and between the 5- to 6-year-olds and the 9- to 10-year-olds (b = -12.7, SE = 3.72, $t(57)$ = -5.832, p < .001). No significant interaction of increase in search attempts remaining over chain positions was found between the 7- to 8-year-olds and the 9- to 10-year-olds (b = 1, SE = 3.72, $t(57)$ = 0.269, p = .961).

Intentional Condition:

No significant main effect of centred chain position was found in the Intentional condition (chain included as a random variable) (b = 1.80, SE = 1.843, $t(74)$ = 0.976, p = .332).

In the Intentional condition, there was no significant difference between search attempts remaining between the 5- to 6-year-olds and the 7- to 8-year-olds (b = 0, SE = 1.37, $t(12)$ = 0, p = 1). The number of search attempts remaining was significantly lower for the 7- to 8-year-olds when compared to the 9- to 10-year-olds (b = -4.4, SE = 1.37,

$t(12) = -3.207$, $p = .019$), and significantly lower for the 5- to 6-year-olds when compared to the 9- to 10-year-olds ($b = -4.4$, $SE = 1.37$, $t(12) = -3.207$, $p = .019$).

In the Intentional condition, no significant interaction of increase in search attempts remaining over chain positions was found between the 5- to 6-year-olds and the 7- to 8-year-olds ($b = 1.3$, $SE = 2.61$, $t(57) = 0.499$, $p = .872$). A significant interaction of increase in search attempts remaining over chain positions was found between the 5- to 6-year-olds and the 9- to 10-year-olds ($b = -12.1$, $SE = 2.61$, $t(57) = -4.641$, $p < .001$), and between the 7- to 8-year-olds and the 9- to 10-year-olds ($b = -13.4$, $SE = 2.61$, $t(57) = -5.140$, $p < .001$).

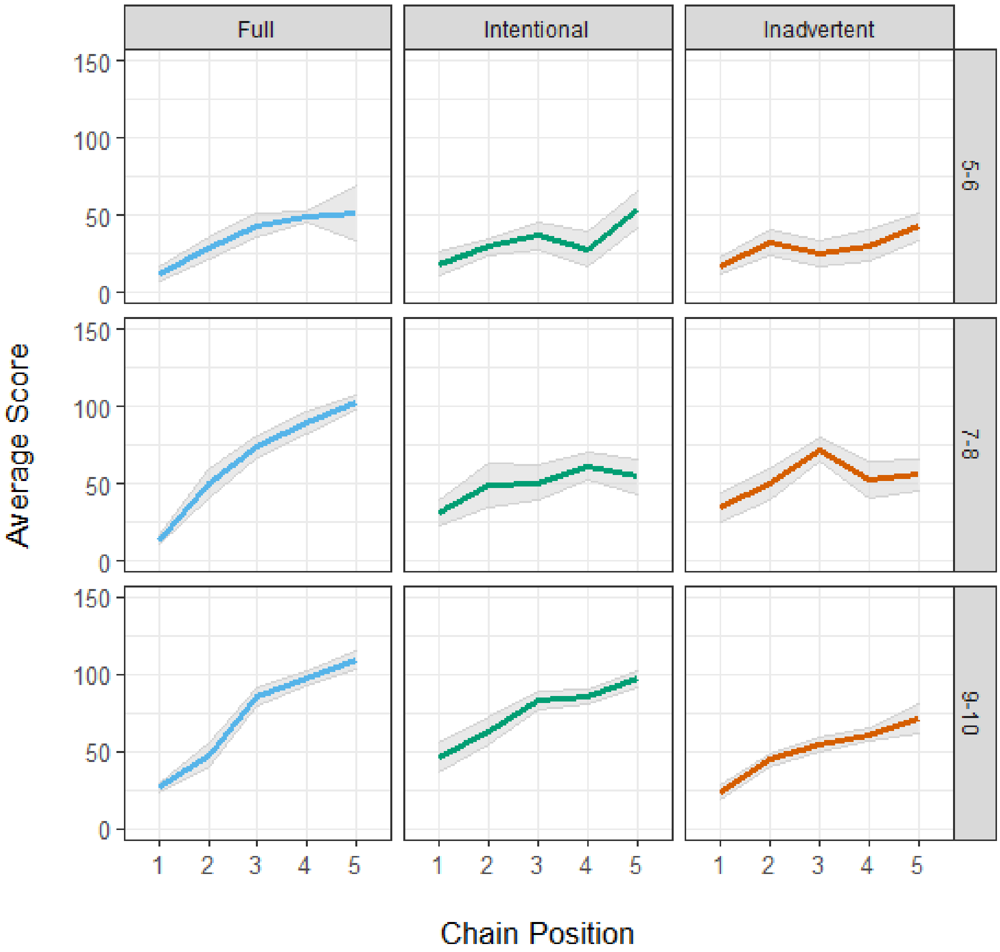
Inadvertent Condition:

No significant main effect of centred chain position was found in the Inadvertent condition ($b = -0.2$, $SE = 0.683$, $t(69) = -0.293$, $p = .771$). In the Inadvertent condition, no significant differences in search attempts remaining were found between the 5- to 6-year-olds and the 7- to 8-year-olds ($b = -0.56$, $SE = 0.273$, $t(69) = -2.049$, $p = .108$), between the 5- to 6-year-olds and the 9- to 10-year-olds ($b = 0$, $SE = 0.273$, $t(69) = 0$, $p = 1$), or between the 7- to 8-year-olds and the 9- to 10-year-olds ($b = 0.56$, $SE = 0.273$, $t(69) = 2.049$, $p = .108$).

In the Inadvertent condition, no significant interaction was found between the 5- to 6-year-olds and the 7- to 8-year-olds ($b = 1.0$, $SE = 0.966$, $t(69) = 1.035$, $p = .557$), between the 5- to 6-year-olds and the 9- to 10-year-olds ($b = -0.4$, $SE = 0.966$, $t(69) = -0.414$, $p = .910$), or between the 7- to 8-year-olds and the 9- to 10-year-olds ($b = -1.4$, $SE = 0.966$, $t(69) = -1.449$, $p = .322$),

Figure 25

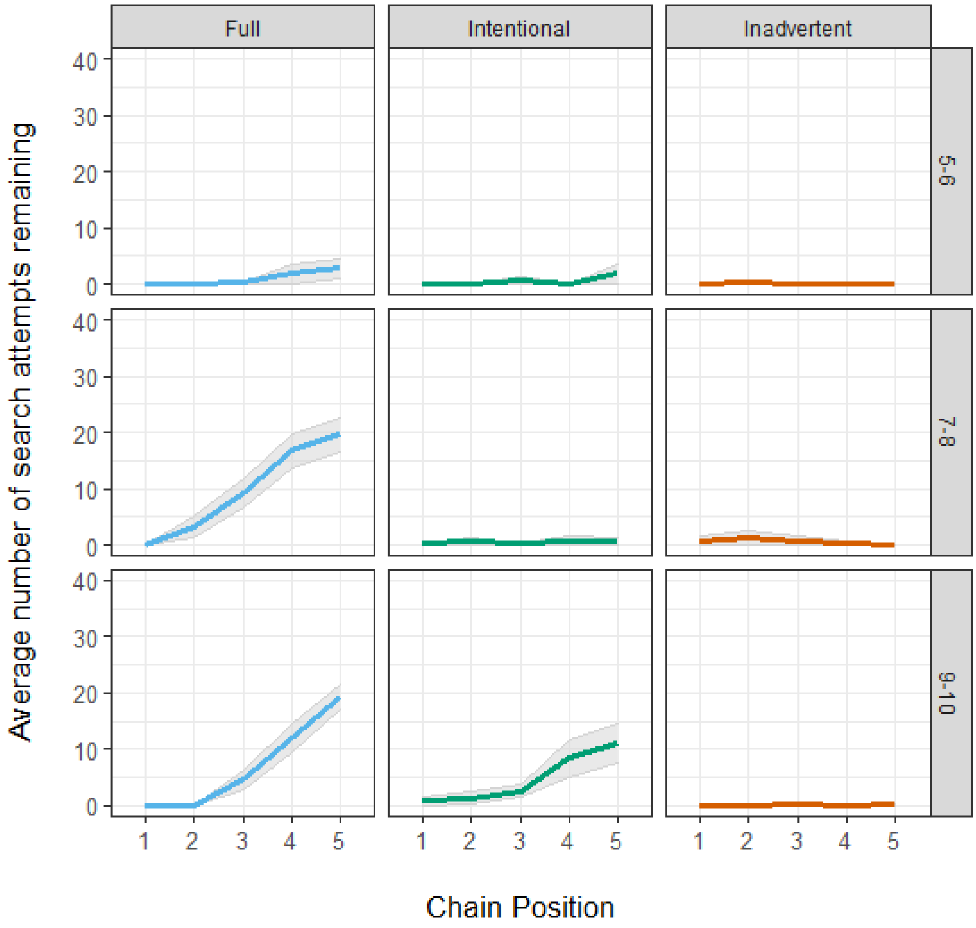
Mean score at each chain position for each condition and age category.



Note. Error areas represent +/- 1 standard error.

Figure 26

Mean search attempts remaining at each chain position for each condition and age category.



Note. Error areas represent +/- 1 standard error.

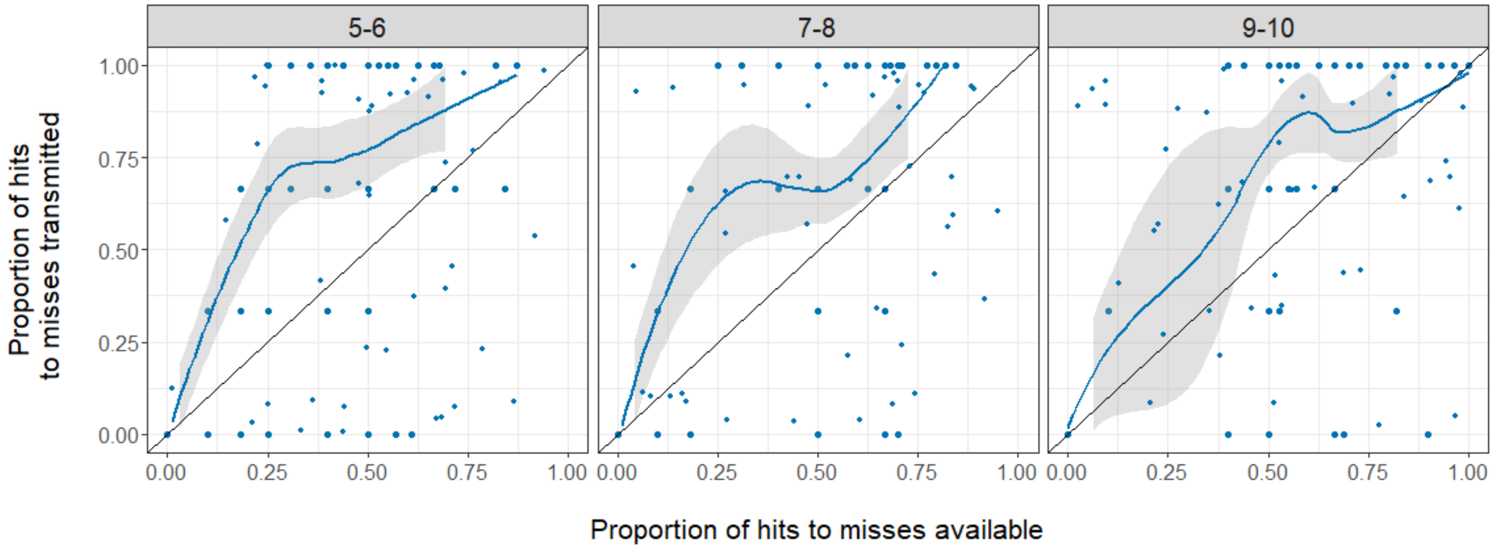
4.4.3 Proportion of hits transmitted in each age category

Paired t-tests were used to compare the proportion of hits available with the proportion of hits transmitted in each age category. In the 5- to 6-year-olds, the proportion of hits transmitted was significantly greater than the proportion of hits available ($t(119) = 7.660$, $p < .001$). In the 7- to 8-year-olds, the proportion of hits transmitted was also significantly greater than the proportion of hits available ($t(124) = 4.766$, $p < .001$). In the 9- to 10-year-olds, the proportion of hits transmitted was also significantly greater than the proportion of hits available ($t(124) = 6.933$, $p < .001$).

We performed a linear model comparing the proportion of hits to misses transmitted across age groups, using the proportion of hits to misses available (centred) as an interaction variable. It was found that 5- to 6-year-olds transmitted a significantly higher proportion of hits to misses compared with the proportion of hits to misses available than 7- to 8-year-olds ($b = 11.03$, $SE = 4.23$, $t(364) = 2.606$, $p = .026$). This difference was not significant between the 5- to 6-year-olds and 9- to 10-year-olds ($b = 7.04$, $SE = 4.56$, $t(364) = 1.544$, $p = .272$) or between the 7- to 8-year-olds and 9- to 10-year-olds ($b = -3.99$, $SE = 4.28$, $t(364) = -0.934$, $p = .619$) (Figure 27).

Figure 27

Proportion of hits transmitted compared with the proportion of hits available.



Note. *Error areas represent +/- 1 standard error.*

4.4.4 The proportion of each tile position selected for transmission in each condition and age category

The data used in this analysis only included information that was produced following a search that uncovered one or more complete targets. Only information generated from complete targets (i.e., no partial targets or misses) was used. The proportion of each tile position selected for transmission in each condition and each age category can be seen in Figure 28.

The positions sent within each ship were analysed to determine whether information senders sent any particular tile position more frequently. The data for this analysis only included tile positions sent from complete targets (out of 750 targets available across all chains, chain positions and grids in each condition for each age category: 5- to 6-year-olds: Full condition, n = 39; Intentional condition, n = 36, Inadvertent condition n = 20; 7- to 8-year-olds: Full condition, n = 116; Intentional condition, n = 61, Inadvertent condition n = 72; 9- to 10-year-olds: Full condition, n = 125; Intentional condition, n = 118, Inadvertent condition n = 59).

To explore whether central tiles were transmitted more often relative to other target segment positions, we first compared the proportion of each target segment selection to a chance level selection. This analysis takes into account that there is only one central segment and eight edge segments in each target, and therefore this chance level was set to 0.11. Binomial tests were then carried out to determine which tile positions were sent significantly more or less than chance.

Pearson's Chi-Square tests were used to compare the distribution of transmitted segments sent in the Intentional condition in each age category to a uniform distribution. It was found that the distribution was not significantly different from a uniform distribution in the 5- to 6-year-old age category ($x(8)$ = 6.234, p = .621, n = 20) or the 7- to 8-year-old age category ($x(8)$ = 6.267, p = .617, n = 72). A significant difference was found in the 9- to 10-year-old age category ($x(8)$ = 21.726, p = .005, n = 59).

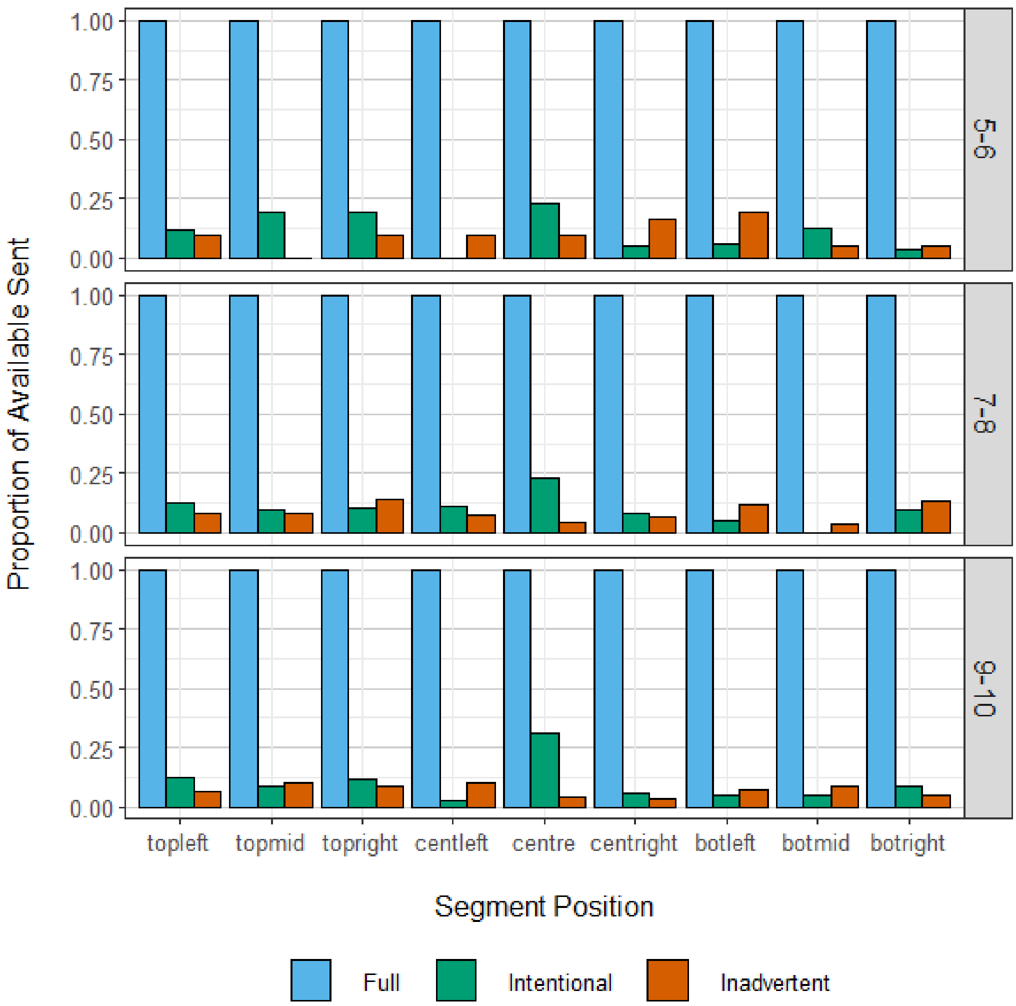
Pearson's Chi-Square tests were used to compare the distribution of transmitted segments sent in the Inadvertent condition in each age category with a uniform distribution. The distribution in all age categories was found to be not significantly different from a uniform distribution (5- to 6-year-olds: ($x(8) = 1.922$, $p = .983$, $n = 36$); 7- to 8-year-olds: $x(8) = 2.655$, $p = .954$, $n = 61$; 9- to 10-year-olds : $x(8) = 1.228$, $p = .996$, $n = 118$).

Because of the low numbers of complete ships uncovered, binomial tests were used to compare the proportion of centre tiles sent with the proportion of edges (non-central target tiles) sent. Again, this analysis takes into account that there is only one centre tile in each target, and eight edge tiles in each target. The proportion of centre tiles sent by the 5- to 6-year-olds was significantly different from the number of edges sent ($x^2(1) = 5.951$, $p = .015$). This comparison was also significantly different in both other age categories (7- to 8-year-olds: $x^2(1) = 11.17$, $p < .001$; 9- to 10-year-olds: $x^2(1) = 58.663$, $p < .001$).

Chi-Square tests were also used to compare the proportion of centre tiles sent between age categories. The proportion of centre tiles sent was not significantly different between the 5- to 6-year-olds and the 7- to 8-year-olds ($x^2(1) = 11.88$ $p = 1$), between the 7- to 8-year-olds and 9- to 10-year-olds categories ($x^2(1) = 0.796$, $p = .372$) or between the 5- to 6-year-olds and 9- to 10-year-olds ($x^2(1) = 0.182$, $p = .670$).

Figure 28

The mean proportion of each target segment selected for transmission in each condition. Split by age category.



Note. This graph incudes data from only completely uncovered targets. This graph therefore does not include any data from partially uncovered targets. X-axis values correspond to particular target segments. For example, “topleft” refers to the top left segment in a target. ‘centre’ in shortened to ‘cent’, and ‘bottom’ is shortened to ‘bot’.

4.5 Discussion

The overarching aim of the previous (Chapter 3) and current studies was to assess whether cumulative culture is facilitated by intentional knowledge transmission from the cultural parent to the cultural offspring, relative to circumstances where knowledge from a cultural parent is acquired incidentally from their performance. In the previous study (Chapter 3), we found that adults successfully accumulated information using intentionally transmitted subsets of the previous participant's search. This accumulation was just as successful in circumstances where participants had full access to the previous participant's search history. In contrast, in circumstances where participants had access to an inadvertent subset of information, accumulation was less successful. However, while this study supports theories claiming cumulative culture relies on distinctively human cognitive mechanisms, adults' information transmissions were only speculatively attributed to these mechanisms, and more specifically, an understanding of other's mind states.

To add weight to claims of human distinctiveness, the current study aimed to explore the developmental trajectory of the capacity for intentional information transmission. Given that children begin to develop a suite of cognitive mechanisms that may facilitate cumulative culture at around 4 years of age, and continue to develop these throughout childhood (Chapter 1), we expected to find that chains of younger children would be unable to accumulate information over generations due to the underdeveloped state of these mechanisms. As such, we expected to find an age-related increase in the ability to select information that facilitates cumulative culture at the population level, consistent with the developmental trajectory of capacities like mental state understanding.

We found that task scores increased over learner generations for the 5- to 6-year-olds, however, all conditions increased at the same rate. This suggests that 5- to 6-year-olds did not fully benefit from the information available to them (since there was no apparent advantage even for full, over incomplete, information).

We also found an increase in task scores over generations in the 7- to 8-year-olds. However, in this age group, scores in the Intentional condition and scores in the Inadvertent condition increased at the same rate, but the increase was significantly greater in the Full condition. This suggests that lifting the bottleneck on transmission meant that children were able to use more information to their advantage, compared with younger children. However, it also suggests that children in this age group were either unable to select the most beneficial information for transmission, or were unable to derive the intended benefit from this, since there was no advantage for the intentional, over inadvertent, condition. Having fewer information points hindered their performance regardless.

In the 9- to 10-year-old age group, we found the same increase in task scores over generations of learners. Chains in the Full condition showed a greater increase in scores than chains in the Intentional condition. However, both of these conditions increased at a significantly greater rate than the Inadvertent condition. This result confirms that intentionally selected information offers benefits over and above a randomly produced sample.

We found a very similar pattern of results with regards to the search attempts remaining following a search. The 5- to 6-year-old chains consistently failed to conserve any search attempts in any of the conditions, and 7- to 8-year-old chains only conserved search attempts in the Full condition, with the number of search attempts remaining increasing over the course of the chains. The 9- to 10-year-old chains, however, increased the number of search attempts remaining over the course of chains in both the Full and Intentional conditions. This implies that children not only became gradually better at accumulating information with age, but were also able to use that information with a greater deal of efficiency (i.e., using fewer search attempts).

The pattern of results found in the 9- to 10-year-old age group matches the pattern found in the data collected from the adult sample in Chapter 3. This suggests that children at this age are, at least approaching, the level of cognitive sophistication required to make complex judgements about what information can be useful to others. This allowed participants to anticipate what information the cultural offspring may need, and in turn make deliberate choices specifically to fulfil that need. The fact that this pattern of results does not emerge until around 9 to 10 years of age suggests that the capacities argued to support intentional information sending in really young children, such as mental state understanding, are not sophisticated enough to apply to scenarios like this until much later in development.

We conducted an exploratory analysis of transmission behaviour in the Intentional condition only. This revealed that children of all age groups selected a disproportionately high number of hits for transmission. Despite the stark differences between age groups in their accumulation of information, we found no significant differences between age groups in this analysis. Information about hits (compared with misses) is much more useful for the cultural offspring, who holds the same goal to score as many points as possible. By selecting the same hit tile, the cultural offspring is guaranteed to gain one point, and they also have a greater chance of locating the adjacent target tiles because of their clustered layout. Conversely, sending a miss location only slightly increases the chance of the receiver gaining a point, compared with sending no information at all. The finding that even very young children preferred to transmit information about a target when it was available suggests that this bias, also seen in the adult participants, may not require a complex understanding of the needs of the learner.

It is perhaps surprising, despite participants in the Intentional condition having a clear bias for transmitting hit information across all age categories, that we still found striking age effects in score increases. These biases clearly allowed cultural offspring in the 9- to 10-year-old age group to perform extremely well with a very limited information regarding their predecessor’s search. However, the same could not be said for the younger age groups. It is possible that there was something intrinsically strategic about the specific locations of the hits being sent in the 9- to 10-year-old age group. Therefore, we evaluated the strategic placement of hits transmitted by each age category to assess whether something about their specific placement was more valuable.

In the adult study, we found a preferential transmission of central target tiles, which presumably aided the cultural offspring’s ability to locate all of the adjacent tiles while retaining as many of their search attempts as possible. In contrast, in the current study we found that when selecting information to send from a completed target, the 5- to 6-year-olds did not send significantly more central target tiles than any other segment.

However, in both the 7- to 8-year-old and 9- to 10-year-old age groups, we found a significant preference for selecting centre tiles to transmit. While we found no difference between the age groups in terms of the proportion of centre tiles transmitted overall, this preference for sending centre tiles and avoidance of sending other tile positions indicates a shift in the older age groups that is consistent with their ability to increase their scores, and decrease the number of search attempts needed to locate all of the targets. This result provides some clarity over the discrepancy between score increase and the preferential selection of hits over misses. Of course, given that the sample size was relatively small, and the results were in the expected direction, it may be possible, with more data, to show that 9- to 10-year-olds were performing much better than was shown here.

It could perhaps be argued that the 5- to 6-year-olds uncovered so few ships that there were too few data points to capture a preference. If this was the case, it could be possible that younger children's performance suffered from an inability to fully benefit from the information they received, rather than a limited capacity to send more useful information. If this were the case, this would mean that children's ability to assess others' needs might develop earlier than their ability to fully appreciate the value of information that has been provided by others. As a result, future research would benefit from exploring how children of all age categories would strategically select information in a task where all targets were revealed following the search. Such a task could potentially use as stimuli the information generated in the current task to ensure all children had equivalent opportunity to select information from the same sample, for a direct comparison of the extent of their information-sending biases.

In this study, we attempted to provide further support for mental state understanding as a specific mechanism that could be facilitative of cumulative culture. Before starting the experimental task, all children were tested using an adapted mental state understanding test deemed suitable for testing older children. This task, ('*the Yanny task*'), assessed both first-order and second-order mental state understanding. In both First- and Second-order conditions, we found an increase in the proportion of correct trials, over ages, with older children in both conditions giving more correct answers than younger children. However, we also found that children generally gave more correct answers in the Second-order condition, which was very unexpected given that it was the most difficult of the two conditions. Furthermore, practise effects were found in both conditions. Because neither the expected age or stage effects were found in the task (likely due to ceiling effects), this analysis was not used to support the role of mental state understanding in any of the further analyses. Furthermore, during testing, a number of children commented on the fact that they were utilizing the gaze of the Yanny character and facial expressions of the other characters in order to help them complete the sentences. While this is clearly a suitable way for children to score highly in the task, it does not necessarily indicate that children were using mental state understanding over and above forming associations involving basic gaze following and facial expression labelling. We concluded that much more research is required to confirm that this is a suitable test of mental state understanding in older children, and we therefore did not use the data from this task to support any other findings.

Of course, it is not the case that this study, without, or even with, positive evidence from additional correlations with the development of mental state understanding, can pinpoint a specific cognitive mechanism that is responsible for the result. While there is a logical argument for mental state understanding, it can only be stated with certainty that the experimental task was cognitively challenging for participants, such that only older children could produce results that mirror those in the adult study (Chapter 3). A host of cognitive mechanisms develop at around the same age as mental state understanding, and discounting any of them as responsible for the result we found in this study would be premature. Batteries testing for other capacities such as executive functions, particularly those testing for both inhibition and working memory are found to capture a similar developmental trajectory as mental state understanding (Carlson et al., 2002; Carlson & Moses, 2001), with studies finding a correlation between the development of executive functions and mental state understanding in false belief understanding (Devine & Hughes, 2016). The development of language has also been correlated with the development of mental state understanding (Astington & Jenkins, 1999), with some accounts also suggesting that this close correlation indicates a causal relationship (De Villiers, 2007; De Villiers & De Villiers, 2014). As with executive function abilities, language development could easily facilitate cumulative culture in this task.

While it was not necessary to use language explicitly in the task, having more advanced linguistic thinking may have improved older children's ability to think and reason about the task, particularly because they may be more experienced at communicating with others. While it is not possible to confidently state that mental state understanding is a facilitator, it is certain that the results are not due to basic biases or preferences – that is, what participants are doing in this task appears not to be trivial, and hence may be likely to be distinctively human.

Interestingly, the age at which intentional, and even full information becomes useful for the learner in this task, coincides with the age at which higher-order capacities for mental state understanding emerge in children over the age of 4 years. As discussed in the introduction to this chapter (Section 4.2), and in Chapter 1 (section 1.6), mental state acquisition at 4 years of age is not the peak of its development. Rather, it continues to develop in increasing orders of complexity, such that older children and adults can reason about multiple layers of beliefs held by different people. Holding a second-order capacity for mental states may be especially beneficial in this task from the cultural parent's perspective, because it would allow them to assess the cultural offspring's belief about their (the cultural parent's) belief that transmitting a central target segment is more useful than transmitting any other target segment. Furthermore, it may allow an increased understanding of how others are viewing the search space. For example, giving cultural offspring the ability to reason that cultural parents may be transmitting an edge segment because they do not have a centre tile to send, rather than assuming that the cultural parent has opted to change their transmission strategy. In alignment with the results found in this task, the capacity for second-order mental state understanding and perspective-taking is typically found to develop in children at around the age of 7 years (Perner & Wimmer, 1985; Salatas & Flavell, 1976), and there is evidence to suggest that development of even higher orders (i.e., third-order and beyond) continues into adulthood (Dumontheil et al., 2010; Valle et al., 2015). While the claim that mental state understanding and related capacities facilitate cumulative culture in this task must remain speculative, it is not unreasonable to suggest that the developmental trajectory found in this task may be related to the trajectory for higher-order mental state understanding.

Overall, our findings suggest a developmental trajectory in refining the intentional selection of information for the benefit of others. Older children (9 to 10 years old) were able to select a very small subset of information that made it possible for later generations to retain the value of a higher volume of information, compared what was found based on transmission of randomly-selected subsets. As a result, cumulative cultural evolution was facilitated when information sending was intentional. Because of the age differences found in this task, particularly in sending and interpreting intentional information, we suggest that the biases found in the intentional information condition in the adult battleship task (i.e., sending centre tiles) required a higher order of cognitive mechanisms. However, further research is required to determine whether mental state understanding, specifically, is at work here. Overall, in demonstrating cumulative culture as a result of successful teaching, consistent with an age increase, we add weight to the idea that sending strategically useful information is distinctive in humans.

Chapter 5: The effect of cultural offspring feedback on cumulative cultural evolution in a grid search task

5.1 Abstract

In a previous study, we found that intentional information transmission from cultural parents facilitated the accumulation of beneficial information over generations (Chapter 2). Due to humans' unique propensity for intentional information sharing, we concluded that this (relative to the inadvertent information transmission found in non-humans) may be a key facilitator of uniquely human cumulative culture. However, given that human teaching is often interactive, we followed up this study by introducing bidirectional information sharing, whereby successors (cultural offsprings, *'offsprings'*) could transmit feedback to predecessors (cultural parents, '*parents'*). Participants completed a grid search task whereby they were required to search for hidden targets using a limited number of search attempts. Information about participants' grid search was communicated between dyads. The cultural parent was required to transmit information that they believed would assist the cultural offspring in finding as many targets as possible. The cultural parent always selected the information to be sent to the cultural offspring, whereas feedback from the cultural offspring varied across conditions. The feedback from the cultural offspring either consisted of the cultural offspring's full search attempt, or subset of the cultural offspring's full search attempt, which was either selected by the cultural offspring, or consisted of a random selection from the cultural offspring's search attempt. We also included a condition whereby no feedback was sent from the cultural offspring to the cultural parent. Chains in all conditions increased their scores at the same rate, indicating that in this context, feedback from the cultural offsprings did not affect the accumulation of beneficial information over and above what can be achieved when intentional information is transmitted from the cultural parent.

5.2 Introduction

The current study investigates whether feedback from a cultural offspring can support intentional information transmission from a cultural parent in such a way that cumulative culture is increased beyond what would be possible with a unidirectional communicative structure. In an earlier study (Chapter 3), we explored the facilitative effect of unidirectional intentional information sending, relative to inadvertent information sending, on cumulative cultural evolution in chains of individuals. We found that, in circumstances where there is a bottleneck in transmission (i.e., only a subset of knowledge can be transferred), accumulation of beneficial information still occurs, but only when the transmitted information was intentionally selected by the cultural parent to help the cultural offspring uncover as many rewards as possible. From this result, we know that cultural parents make informed decisions about what to transmit, and that these decisions allow an increased accumulation of benefits. However, we don't yet know anything about the role of the cultural offspring, other than their ability to accurately interpret information from someone else. It is possible that the cognitive abilities of both the cultural parent *and* cultural offspring facilitate distinctively human cumulative culture. Given that we know about the cognitive requirements of the cultural parent in this context, the current study aims to assess whether the cognitive requirements of the cultural offspring matter. Specifically, whether judicious decisions made by cultural offsprings about what information about their knowledge to give to a cultural parent offers benefits over and above those found in Chapter 3.

In most cases of synchronous teaching, cultural parents do not communicate unidirectionally to their cultural offsprings. Feedback from the cultural offspring may assist the cultural parent in fine-tuning the information they provide, such that it becomes more tailored to the needs of the cultural offspring. While the literature supporting this claim is sparse, there is some evidence from empirical research that we believe provides a strong rationale for this expectation. In a task that required participants to tie a variety of knots, Caldwell et al. (2018) found that conditions in which participants were allowed to interact with a teacher (synchronous information transmission) resulted in greater success for the learner than conditions in which participants only had access to pictures of the intermediate steps for knot completion (asynchronous information transmission). In this study, the intermediate steps condition can be seen as a non-interactive teaching condition, since these materials had been produced for the express purpose of guiding a naïve learner through the process. The synchronous, interactive teaching condition did facilitate success for the learner, relative to this non-interactive condition, however only in cases where the knot to produce was deemed relatively difficult. Therefore, this study suggests that interactive teaching may be particularly supportive for hard-to-learn information.

A further area of research that we believe bolsters our reasoning that cultural offspring feedback can facilitate effective teaching is the existing literature surrounding '*overhearers'* in conversational exchanges. These tasks involve interactions between members of a conversational dyad, but focus primarily on the understanding of an external participant (the overhearer), not directly involved in the interaction, who hears part or all of the exchange. As such, overhearers are privy only to information tailored in response to feedback from someone else. The information may therefore be less readily interpretable to the overhearer, despite the fact that they are exposed to identical information content. Overhearers, relative to addressees, have been found to have a reduced ability to match abstract objects to the speaker's descriptions of those objects (Clark & Wilkes-Gibbs, 1986), and retell stories (Schober & Clark, 1989). Therefore, there is a clear advantage to cultural offsprings who can give feedback to cultural parents about their own performance, or communication failures, or gaps in understanding, relative to overhearers who cannot give feedback.

The above studies suggest that interactive teaching can support more accurate transmission of information. However, the intricacies of cultural offspring feedback have not yet been explored in relation to cumulative cultural evolution. Particularly pertinent from the perspective of this thesis, the question remains as to whether the cognitive abilities of the cultural offspring are relevant in the production of feedback designed to help optimise information transmission from the cultural parent, and as a result, the cultural offspring's own knowledge. Alternatively, it may be possible that having access, as the cultural parent, to any information from the cultural offspring's performance (i.e., not necessarily intentionally produced by the offspring) is all that is required to tailor information adequately enough without assistance from the cultural offspring's cognitive capacity. Inadvertent cues like these are likely to be produced by all species. As such, information about the cultural offspring's knowledge and understanding can be indicated by their performance, for example, if a rat is not able to release food in a puzzle box by pressing the lever, then it would be fair to assume that the rat has not made the causal connection between pressing the lever and the release of food. If it is the case that these inadvertent cues are enough to inform the cultural parent, then it is unlikely that the cognitive capacities of cultural offsprings are relevant to the distinctiveness of cumulative culture in humans. If so, this would contrast with the findings from in the previous chapter, which established that the cognitive capacities of cultural parents certainly do appear to be relevant for understanding human cumulative culture.

While cultural offspring *can* passively produce information about their success to the cultural parent without any deliberate transmission of their own knowledge, to produce information that functions to signal to the cultural parent their level of competence, it may be the case that active, intentional offspring feedback offers benefits over and above this. We know that human adults are in principle capable of this, but here we ask whether this makes a difference to the information being sent from the cultural parent, or if passive cues from cultural offspring are just as effective.

We attempted to apply a bidirectional communication structure to the 'Battleships' paradigm outlined in Chapter 3, to assess whether a bidirectional information transfer increases the accumulation of information over generations. To do this we introduced a transmission channel allowing communication from cultural parent to cultural offspring (downward transmission), but also cultural offspring to cultural parent (upward transmission). Each pair of participants engaged in several interactions before the parent participant was replaced, with the former cultural offspring taking on the role of cultural parent, and a naïve participant being introduced as the new cultural offspring (See Figure 29).

In this task, all transmissions from the cultural parent in the dyad were made up of three selections from their search, all intentionally selected by the cultural parent. Therefore, in all conditions in this task, information from the cultural parent to the cultural offspring was exactly as per the information sent in the intentional condition of the previous task (Chapter 3). This is because we already determined the effectiveness of intentional communication in Chapter 3, and this study only aimed to explore conditions that would improve intentional communication from the cultural parent.

The feedback response from the cultural offspring was manipulated to provide either:

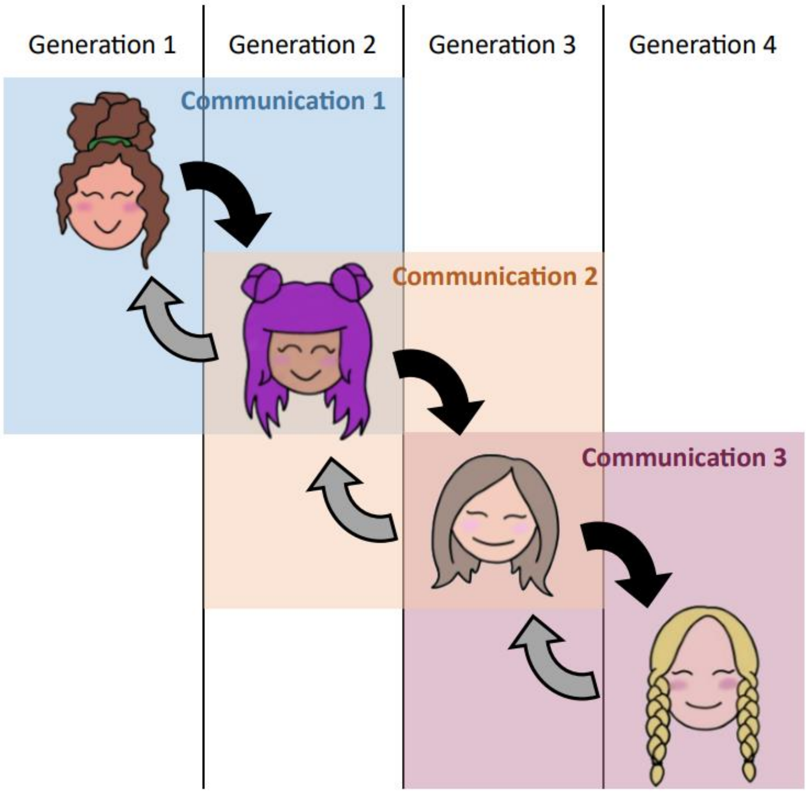
- **An Intentional subset** of the cultural offspring's search selected by the cultural offspring themselves.
- **An Inadvertent subset** of the cultural offspring's search randomly generated from the cultural offspring's selections.
- **Full information** from the cultural offspring's search (maximum feedback comparison).
- **Null information**, i.e., no feedback was given from cultural offspring to cultural parent (no feedback baseline).

Our main interest was whether this dyadic communication facilitated cumulative cultural evolution. We expected to see cumulative cultural evolution in every condition because the results of the unidirectional battleships study should hold true for this study too (given that all cultural parents send intentional information). However, there are two possible outcomes for the effect of offspring feedback – either the feedback is useful to the cultural parent but without any added advantage of this being intentionally produced, or offspring feedback is useful but particularly so if it is intentionally produced. The former would suggest that offspring feedback might facilitate cumulative cultural evolution, but the cognitive capacities of the offspring may not be relevant in this context, i.e., there would be nothing to suggest that this might be part of the explanation for the existence of uniquely human cumulative culture. However, if the latter was true, this would suggest that the offspring's cognitive capacities *are* relevant, and may go together with the cultural parent's cognitive abilities as part of the explanation for distinctively human culture.

In the event that we did find an effect of intentional cultural offspring feedback, we planned a further exploratory analysis assessing the strategic transmission of specific target segments. This is analogous to the analysis performed in Chapter 3 which revealed that cultural parents had a bias for transmitting central target segments over edges. The previous task used 3x3 targets, meaning that cultural parents had the option to send a centre target segment to indicate the location of a full target. Making this selection was a somewhat obvious *best choice* about which tile would be the most useful for the cultural offspring to know about, and it did allow the cultural offspring to find all of the remaining tiles in that target without incurring loss. If the cultural parent opted to send a different target segment (i.e., an edge square), the cultural offspring was more at risk of failing to anticipate which edge square the cultural parent selected, and therefore selecting non-target grid squares. To explore the effect of cultural offspring feedback and grounding, which might lead to the establishment of arbitrary communicative conventions, we reduced the target to a 2x2 grid size, thereby eliminating any objectively preferential options, allowing the possibility of establishing arbitrary conventions for which tile to use to indicate the presence of a target.

Figure 29

Organisation of participants in the task.



Note. Participants involved in interactions are included together in coloured boxes. Parent participants are placed higher in the communication box and offspring participants are placed lower. Black arrows indicate downward transmission from cultural parent to cultural offspring and grey arrows indicate upward transmission from offspring to parent.

Based on the results of the previous study (Chapter 3), and the evidence supporting a facilitatory effect of offspring feedback, we make the following predictions:

- Participants assigned to later generations in the chains will achieve higher scores than those assigned to earlier generations.
- The ordering of conditions in relation to the total score will be:
  Full Feedback (highest scores) > Intentional Feedback > Inadvertent Feedback > No Feedback (lowest scores).
- There will be an interaction between condition and generation such that any differences in task score found between conditions will be more apparent in later, compared with earlier generations.
- The ordering of the conditions in relation to the generation by which all targets will be found will be:
  Full Feedback (earliest) > Intentional Feedback > Inadvertent Feedback > No Feedback (latest).
- The ordering of the conditions in relation to the number of search attempts required to find all targets will be:

  Full Feedback (fewest misses) > Intentional Feedback > Inadvertent Feedback > No Feedback (highest number of misses).

Our key predictions were pre-registered: https://osf.io/9zp4j/

5.3 Methods

5.3.1 Participants

All participants were recruited as part of a second-year undergraduate lab practical class, and took part on a voluntary basis. Participants arrived to be tested in groups of up to 20. One group per hour was tested, and chains left incomplete after the hour session were either completed in the next session or discarded in the event that no session followed.

Four computers were used to host the training task, and a further six computers were used to run the experimental task. As such, three pairs of participants could be tested at any given time, which meant that multiple chains were running simultaneously.

Two hundred and forty-three adults (199 female) were recruited in total. Participants were given three age bandings as options for reporting age and as such, age distribution was as follows: 224 x 16- to 25-year olds, 14 x 26- to 35-year-olds, 5 x 36- to 45-year-olds. 27 of these participants (all 16-25, 25 females) were excluded from the study due to experimenter or technical error, or due to incomplete chains which had to be discarded. Because of the nature of the recruitment process, it was common to have to end the session part-way through a chain. Incomplete chains could not be revisited at a later date because previous interactive members were no longer available to take part. The final sample consisted of 216 adult participants (175 female, age banding distribution: 197 x 16- to 25-year-olds, 14 x 26- to 35-year-olds, 5 x 36- to 45-year-olds).

The Full and Null conditions each contained 52 participants (13 complete chains), and the Intentional and Inadvertent conditions each contained 56 participants (14 complete chains). Participants were assigned to a condition randomly, and the chain position was constrained by the current state of completion of the chain to which they were recruited. If a participant's data was excluded, this was replaced by a new participant's data before the chain progressed further.

Chains were made up of four participants. We reduced the number of participants from ten (as was the chain length in Chapter 3) for several reasons. The first is because we found an accumulation of information in the Intentional condition at a much earlier chain position than expected in the previous task. Given that the transmission from the cultural parent to the cultural offspring in this task was identical to the Intentional condition in the previous task, we shortened the chains accordingly. The second reason is that we determined it unnecessary to demonstrate cumulative culture on a larger scale, and instead focus on the inherent benefits of learner feedback on a smaller scale. The third reason concerned the logistics of collecting the data. As aforementioned, it was common for a testing session to end part-way through a chain, which could not be completed in a later testing session. Reducing the number of participants in the chain drastically reduced the loss of data as a result of this, and meant that a higher number of chains reached completion.

The study was approved by the University of Stirling General University Ethics Panel (approval reference number: GUEP 470) and every participant gave consent via a Qualtrics (2020) questionnaire before the experiment took place. Qualtrics assigned each participant a random 6-digit number that they could use to withdraw their data should they wish to do so.

5.3.2 The experimental task

This task, including the training task, was run using PsychoPy v1.84.2 (Peirce et al., 2019), on desktop computers running Windows 10. A total of six computers were used for the task simultaneously, and a further three were used to run the training task.

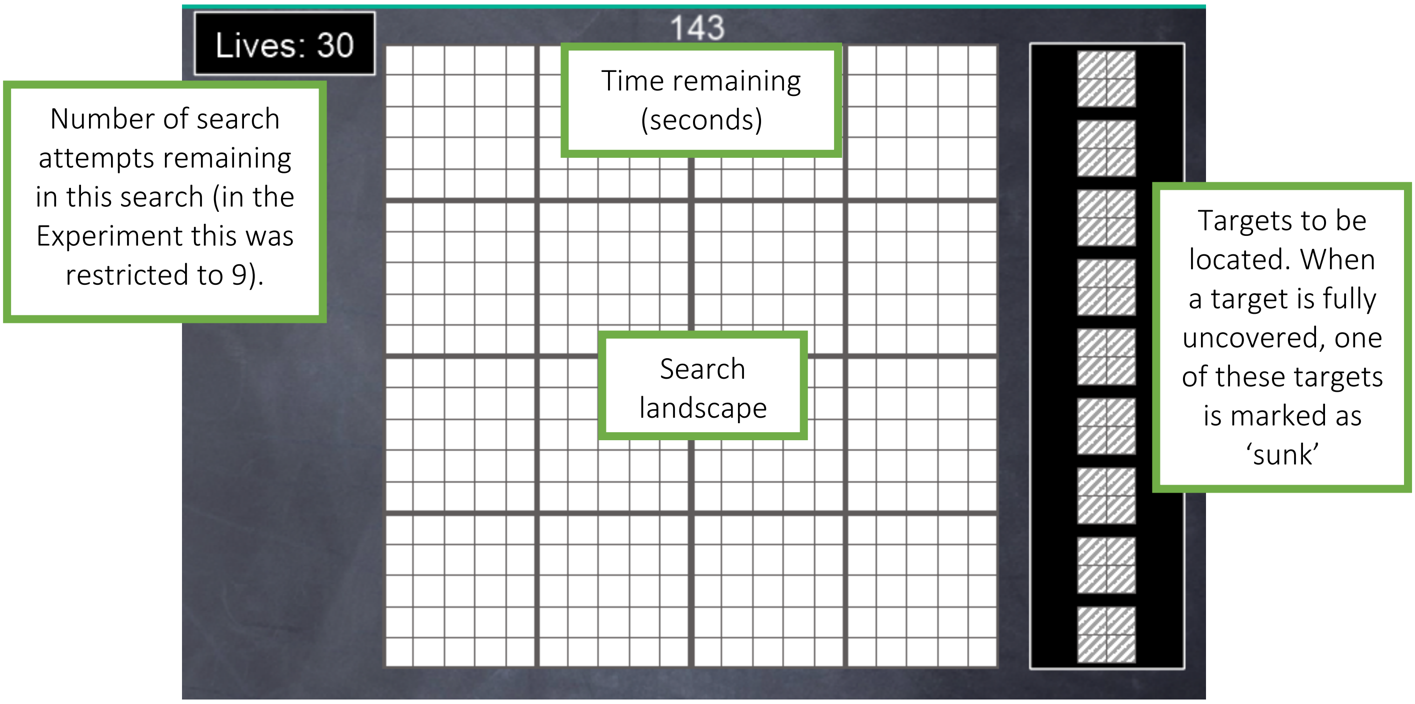
During the experiment, participants played a landscape searching task based on the popular game, '*Battleships'*. Participants were shown three reward landscapes presented as 20x20 grids. Grids were partitioned into 16 sections as a visual aid (Figure 30). Each grid contained nine hidden 2x2 targets to be found. Targets were randomly placed and consisted of four adjoining grid squares. The grid partitions did not affect the placement of targets, and targets could be placed over partitions, in more than one section. Targets were found by clicking the unselected squares on the grid. When a part of a target was located, this selection was coded as a '*hit'*, and was replaced with a white square containing a green pentagon to indicate that one point had been gained. Selection of a non-target grid point was coded as a '*miss'* and was highlighted by white square containing a red circle indicating the loss of one search attempt. The primary goal of all participants was to score as many points as possible over all three grids using as few search attempts as possible.

Participants used one of their nine search attempts for every selection they made that reveals a non-target grid square. In contrast, the participant gained one point for every target segment they found, with the number of search attempts remaining being preserved in the process. This structure was designed to reflect real world relative payoffs, such that a search attempt resulting in a reward was always profitable, while a search attempt resulting in no reward was costly.

The maximum score for each grid was 36 (nine targets each made up of four individual grid squares). The minimum number of selections required was nine (indicative of the number of search attempts allotted per search). In theory, the maximum number of selections per search attempt was 44, should a participant have found 35 of the target grid squares, using all nine search attempts to do so. The participant moved on to the communication segment or next grid once they had located all nine hidden targets, or if they used all nine search attempts before finding all of the targets.

Figure 30

An annotated blank training grid before the participant has made any selections.



Note. The number of search attempts remaining is shown in the top-left corner, the time remaining (number of seconds) is shown above the grid, and the number of targets left to find are shown to the right of the grid. The time remaining is an aesthetic feature to encourage participants to focus on the task only, and if the time remaining reaches 0 seconds, the task does not end, and participants can continue to make selections. Every time a participant uncovers a target, one of the targets shown to the right of the grid is highlighted and remains highlighted until the grid is complete. Participants can click on any of the grid squares to uncover whether it is a hit or a miss. Bold lines separating the grid are a visual aid only and do not imply the locations of any hidden targets. Green annotation boxes were not visible to participants.

Communication

The first participant in each chain received no information at all from a cultural parent. Participants in all other chain positions were given guidance in stages by their immediate predecessor (cultural parent). This guidance was given in three parts over the course of the grid search, such that one information point was given between each search expedition carried out by the cultural offspring. In the current task, we maintained the number of cultural parent transmissions that we allowed in previous tasks (three), however in those previous tasks, this amounted to one transmission per target available to find. This means that in the current task, if the cultural parent uncovered all nine targets, they would only be able to send information from three. The reason we did this was because, to see an effect of cultural offspring feedback, the task needed to be restricted from the cultural parent's end. That is, if the parent could send all of the available information, we would not be able to draw out an effect of the manipulation.

The reward landscape was pre-specified for each grid in each chain of participants, meaning that the results of the cultural parent stayed true for the cultural offspring. cultural parents saw some, all or none of the search attempts made by the cultural offspring. All transmitted information was completely transparent such that participants could see whether the grid square selected was a hit or a miss (characterised in the same way as a selected square but with an orange border to indicate that it had been sent from another participant). This information varied depending on the condition that the chain was assigned, and was applied to all members of the chain apart from the first chain position, who received no information.

- **Null Feedback Condition (control)**: Cultural parents were given no information from the cultural offspring about the progression of their search, and were therefore required to send guidance to the cultural offspring with no feedback. This condition was one of two controls, and was designed to mimic situations in which social information is required, but no access to the cultural offspring's progress is available.

- **Inadvertent Feedback Condition**: Cultural parents received a subset of three information points that had been randomly generated from the cultural offspring's immediately previous search attempt. This condition was designed to reflect inadvertent feedback available from a cultural offspring's performance.
- **Intentional Feedback Condition**: Following their search attempt, cultural offsprings were asked to select three of the uncovered grid points to send to the cultural parent as a means to guide the cultural parent 's downward transmissions. Therefore, in this condition, both cultural parents and cultural offspring transmitted intentional information.
- **Full Feedback Condition**: Cultural parents received all of the selections made by the cultural offspring during the immediately previous search attempt. Therefore, cultural parents in this condition received between nine and 44 grid squares.

5.3.3 Procedure

Training

All participants completed a training session in which they were given all of the instructions for the game including a demonstration of an unselected grid square and grid squares transmitted from a previous participant showing both hits and misses. Instructions given to participants and a run-through of the training task can be seen in Appendix 8. Following the instructions, participants were given three grids identical to that of the real experiment, however, they showed no selected grid squares. These grids allowed completely free exploration with unlimited search attempts, and only progressed to a new grid once all of the targets were found. Participants remained on the training session until there was a free slot in the test stage, or until they had completed a minimum of one full grid and reported that they were confident with the instructions.

The experimental task

The first generation of participants searched all three test grids without any communication phases before taking the role of the cultural parent to assist in the second generation's learning stage. Therefore, generation one had no information about the location or presence of any of the hits or misses in any of the three grids prior to their search. All other generations received one completely transparent (that is, with hit/miss information visible) grid point in each transmission of the communication phase (three in total). The information from cultural parent to cultural offspring was always intentional, and was selected by the cultural parent from the selections they had made available during their own search.

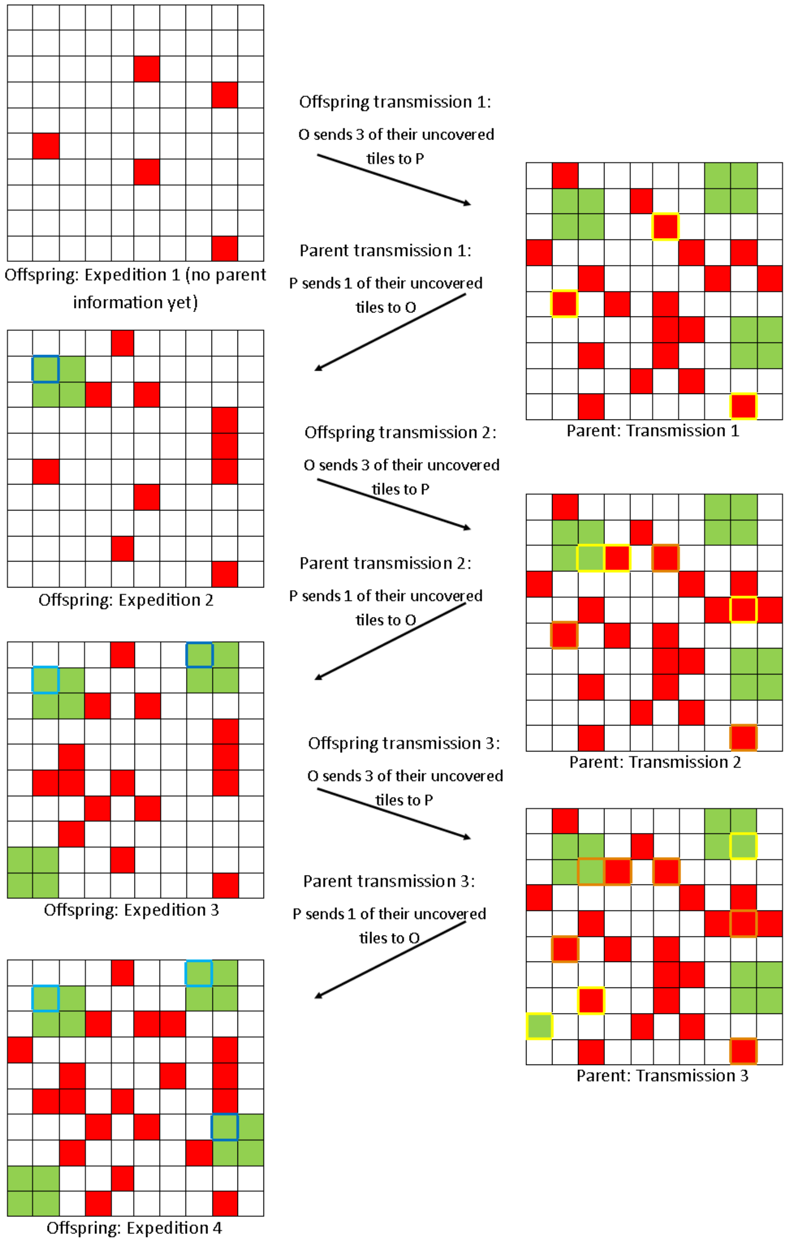
Newly introduced participants completed four '*expeditions'* of each grid, making a minimum of 9 selections (i.e., the number of search attempts) in each expedition, and therefore using a minimum of 36 search attempts for each of the three grids. In the Intentional condition, immediately following each expedition of each grid, cultural offspring were asked to select three of the squares selected during that expedition to send to the cultural parent (the previous generation's cultural offspring). The three squares selected by the cultural offspring were immediately transmitted to the cultural parent who could then respond by transmitting one of their own selections from any of the grid points revealed during their own search of the same grid. Cultural parent transmissions were sent before the cultural offspring continued with the next search expedition. During the cultural offspring's search, cultural parents were able to view the end product of their own grid search, and could make selections to send from the uncovered grid squares (both hits and misses). A timeline of the task, including instructions screens shown to participants and a run-through of the task screens can be seen in Appendix 9, and a scaled-down version of the task to demonstrate the direction of transmission can be seen in Figure 31.

Both cultural parent and cultural offspring were present during searches and transmission stages because the selections were transmitted in real time through a computer link. Participants who were interacting in the same chain were not positioned on computers next to one another, in order to avoid forbidden conferring.

After all three grids had been completed the cultural offspring took the role of cultural parent and a new participant was recruited as the cultural offspring. The former cultural parent was fully debriefed regarding aims and hypotheses and given the opportunity to ask questions.

Figure 31

Scaled-down demonstration of the experimental task.



Note. In the real experimental task, grid sizes were 20x20, however, the size and colours have been edited for visibility in this diagram. Cultural offspring (O) progress is shown on the left of this diagram, and cultural parent (P) information is shown on the right. In this diagram, red-filled grid squares represent uncovered misses and green-filled grid squares represent uncovered hits. A dark blue border represents the most recent cultural parent transmitted information, and a light blue border represents parent information from an earlier transmission. A yellow border represents the most recent cultural offspring transmitted information, and an orange border represents cultural offspring information from an earlier transmission. Arrows represent the direction of transmission. A colour key for the real experimental task can be found in Appendix 9.

5.4 Results

All generalised linear mixed models were performed using (R Core Team, 2020) and lme4 (Bates et al., 2014). Graphs were created using *ggplot 2* performed on R Studio (RStudio Team, 2020), supported by *R version 4.0.2*. The dependant variable in this analysis was score (numeric). Fixed effects are specified before the model output.

Post-hoc comparisons were done using the *emmeans* package (Lenth et al., 2019). Tests were two-tailed, and p-values of < .05 were taken as statistically significant. Emmeans using the Tukey adjustment was used for multiple comparisons.

5.4.1 Mean score increase over generations

Mean scores were calculated using each participant's sum scores for all grids, over all expeditions, across all three completed grids. Mean scores at each chain position can be seen in Figure 32.

The search attempts remaining for each participant were not calculated, due to there being no instance where any participant found all nine targets, thus the number of search attempts remaining was the same for all participants across all conditions.

Linear mixed effects modelling was used to calculate the differences in overall scores using condition and chain position (centred) as fixed effects. Chain was also included as a random variable. A significant main effect of chain position was found (b = 12.307, SE = 1.53, $t(647)$ = 8.038, p < .001, with participants in later generations scoring higher than earlier ones.

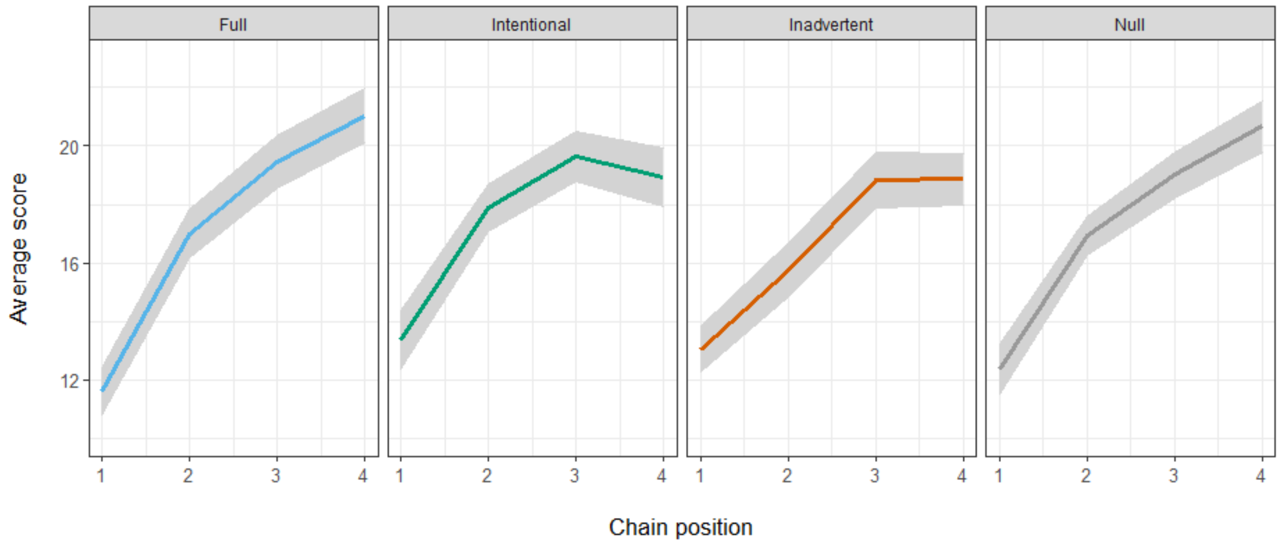
Emmeans comparisons were used to explore the differences between conditions in terms of score increase over generations.

No significant differences were found when comparing the Full and Intentional conditions (b = -0.183, SE = 0.89, $t(50)$ = -0.205, p = .997), or the Full and Inadvertent conditions (b = 0.644, SE = 0.89, $t(50)$ = 0.723, p = .888) A non-significant difference was also found between the Full and Null conditions (b = 0.032, SE = 0.91, $t(50)$ = 0.035, p = 1). There was also no significant differences found between the Intentional and

Inadvertent conditions (b = 0.827, SE = 0.88, $t(50)$ = 0.946, p = .780), the Inadvertent and Null conditions (b = -0.612, SE = 0.89, $t(50)$ = -0.687, p = .902), or the Intentional and Null conditions (b = 0.215, SE = 0.89, $t(50)$ = 0.241, p = .995).

Figure 32

Mean score increase at each chain position, split by condition.



Note. Error areas represent +/- 1 standard deviation.

5.5 Discussion

The previous chapters in this thesis have found that cumulative cultural evolution is enhanced by intentional unidirectional information transfer. The general finding from the earlier study was that (adult) cultural offsprings are able to derive much more value from a subset of information that has been intentionally produced, compared with a subset that was produced inadvertently. However, given that some human teaching also involves intentional upward transmission from the learner, and given that this does probably does not occur in animal teaching, the current study followed on from the previous experiments by introducing bidirectional information transfer. This enabled the cultural offspring to transmit some feedback on their performance to the cultural parent over three transmissions. The main aim of this study was to assess whether the supporting role of intentional knowledge transmission in cumulative cultural evolution could be reinforced by the addition of intermittent upward transmissions from the cultural offspring to the cultural parent. Following a similar logic to the previous studies in this thesis, the amount of information transmitted was limited, to allow for comparison between intentionally selected versus randomly selected information. While the downward transmissions from the cultural parent to the cultural offspring were always intentional, the upward transmissions from the cultural offspring to the cultural parent differed depending on the condition being tested in the chain. We were particularly interested in how the two limited feedback conditions (Intentional and Inadvertent) compared to each other, and how they compared to a Full and a Null condition, in which cultural parents received all or none of the information generated by the cultural offspring's activity respectively.

Given the score increase found in the Intentional condition in the unidirectional task (Chapter 3) and given that all downward transmission in this task was also intentional, we expected to find an increase across all conditions. However, we expected to find a steeper increase in conditions where cultural parents were given access to full or intentional information from the cultural offspring's search attempt, as compared with either inadvertent information, or no information.

We found a significant score increase over generations in every condition, however, there was no difference in the rate of this increase between any of the conditions. This result supports previous findings that intentional information transfer from the cultural parent to the cultural offspring facilitates cumulative culture, however it does not provide any evidence that feedback from the cultural offspring (revealing information about the outcome of the cultural parent's transmissions), at least in this context, provides any additional facilitative effect. For this reason, we did not perform any additional analysis to assess transmission biases, or alignment of transmission biases within chains.

There are several possible reasons why we didn't find any differences in score increase between conditions. The first is that, very simply, cultural offspring feedback has absolutely no effect, and that accumulation in this task rests solely on intentional information being transmitted from the cultural parent to the cultural offspring. Although we still saw a small accumulation over chains, in most circumstances, cultural parents had a very limited selection of useful information available to send, which promoted the sending of redundant information, quite possibly not even chosen for its potential communicative value (see below).

Redundancy

In the unidirectional battleships task, participants transmitted information from one of two categories – it was either a hit or a miss. All of the information they could potentially choose was new information to the cultural offspring. However, in the current task, information that could be transmitted (in either direction) could be, for the recipient, an unknown hit, an unknown miss, or a hit or miss that had already known as a result of either their own expedition, or in a previous transmission. There is great variation in the value of sending each of these different categories of segment. Furthermore, the sender may or may not be aware of whether they are providing new, or old, information to the recipient.

1. A hit that is novel to the receiver is arguably the most valuable information for downward transmissions, because it's alerting the communicative partner to the location of a target that they were previously unaware of the location of. This is particularly useful information because the cultural offspring is guaranteed to increase their score as a result. Novel information is not useful to the cultural parent because participants can only transmit information from grid squares that have been uncovered on their own grids.
2. It is possible that sending redundant hits bares some value in certain contexts too. For example, to signal to the parent sender of a novel hit that the offspring has found the remainder of that target. Of course, this assumes that the cultural parent needs an indication that that target has been found. If this is the case, sending that hit is a bit less redundant because it is necessary to stop any more redundant segments from that target being sent from the cultural parent. Arguably, these selections are the most valuable for upward transmissions.
3. The value of misses is equal whether they are recently uncovered or from a previous search. In downward transmissions, these selections only serve to indicate to the cultural offspring that there is nothing to be found in that specific grid square, however in some cases, sections made near the edge of the landscape boundary, may provide a limited amount of further information (i.e., if it isn't possible for a 2x2 ship to be located in the space between the miss selection and the boundary). The value that these misses hold are mostly relevant when they're sent from the cultural parent to the cultural offspring, because they give an indication of where not to search. However, misses transmitted from the cultural offspring to the cultural parent indicate more generally that in some circumstances, the cultural offspring has no useful information to send, and so any outstanding targets that the cultural parent knows about should be transmitted.

Methodological limitations

Given that communication from the cultural parent happened between cultural offspring searches, cultural parents' sending of offspring-novel information relied on the cultural parents having further information to send which was still novel to the cultural offspring, even after the cultural offspring's own expedition(s). However, if the cultural offspring located all of the targets found by the cultural parent, the cultural parent would no longer be in a position to provide the most valuable form of information (i.e., unknown hits). This was relevant across all conditions, as it was always likely that cultural offsprings would tend to find the same targets as their cultural parents, given that they were effectively being guided towards them by the information they received from their parents. Therefore, in retrospect, we may have made it difficult to find any differences between conditions. In future research, it might be possible to rectify this by altering the parameters of the task to reduce the possibility of cultural offspring finding all of the rewards discovered by their cultural parent before the cultural parent had completed their downward transmissions.

It may also be possible that search biases can explain why there was so little novel information to send between dyads. In a task using an identical grid layout, Atkinson et al. (2020) found that participants selected certain grid squares (particularly tiles centrally located within the landscape, or bordering the partitioned grid sections) much more frequently than other tiles. If participants in this task has similar search biases, in addition to making similar discoveries as a result of cultural parent transmissions guiding them to certain areas of the grid, it would be a lot less likely for novel targets to emerge.

Proposed adjustments for future research

If we were to find any benefit of intentional information sending from the cultural offspring in this study, this depended on an assumption that the amount and/or quality of cultural offspring feedback did matter. However, this basic assumption was not fulfilled, in that there was no difference found between the Full condition (maximum possible feedback about cultural offspring performance) compared with the Null condition (no information at all about cultural offspring performance). As discussed above, finding a difference between these conditions may have been unlikely due to the parameters of the task. It may be possible to draw out a difference between these conditions with a few parameter changes, such as altering the number of selections given to cultural offspring before they transmit information to the cultural parent. Finding a difference between these two extreme comparison conditions would be essential in order to infer any conclusion regarding the effect of the intentionality of cultural offspring feedback. In the absence of such an effect, we cannot conclude anything regarding whether intentional cultural offspring feedback is any more effective than inadvertent cultural offspring feedback.

Future research may explore the search biases made by participants to assess whether common clustered searching restricted the number of uncovered targets. If the same search biases are found in this task as were found in the study conducted by (Atkinson, Blakey, & Caldwell, 2020), this would explain why participants had very little information to send between them. While increasing the target size may mediate for this to some extent, making some adjustments to the search environment (such as removing partitions in the grid) might also help.

A new direction, and possible avenue for future research would be to change the goal of the task in order to explore contexts in which offspring feedback is beneficial for an accumulation of information. In the current experiment, participants were required to locate specific target points on the grid, and only those target points held value. There was very little value in sending information from unrewarded tiles. However, in a context where the goal is to move from one side of the grid to the other side using a path of connected grid squares while avoiding obstacles, having feedback from an offspring may provide a much richer base for the parent to form meaningful transmissions. Obstacle information would be highly valuable, in its role as something to *avoid*. However, information about unobstructed routes would also be valuable. Furthermore, feedback from the cultural offspring would be highly informative, as an indication of progression, thus revealing not only awareness about the status of the current location, but also providing an indication of what additional information would be relevant (onward route) and what would be now redundant (locations already passed). This information would allow the cultural parent to apply meaningful and relevant advice that might subsequently result in an increased success rate.

Conclusions

Overall, our findings support the premise that intentional information sharing can result in a highly efficient transmission of beneficial information, even in very early generations. However, we found no further facilitative effect when a feedback loop was introduced. This may be due to methodological factors which were required to allow close comparison with the previous study (Chapter 3) but which may have constrained detection of an effect of offspring feedback. As such, the lack of effect in this task does not necessarily indicate that cumulative culture is not supported by information transmission with cultural offspring feedback. Further research is required to address this.

Chapter 6: General Discussion

6.1. Summary of overall aims and hypotheses

This thesis aimed to contribute to the existing literature exploring distinctively human cognitive mechanisms. Specifically, the aim was to provide support that the distinctively human propensity for understanding the mental states of others facilitates distinctively human cumulative cultural evolution.

We explored this concept by altering aspects of traditional lab-based tests of cumulative culture to assess conditions in which intentional knowledge sharing is restricted in volume such that earlier generations have to apply some mental state understanding to make beneficial information more accessible to later generations. We compared this to conditions representing circumstances in which knowledge is transmitted inadvertently (i.e., selected without intention), which is likely to be the kind of transmission that supports social learning in non-humans.

The motivation for exploring the role of distinctively human cognitive mechanisms in cumulative culture came from existing accounts that link them (e.g., Boyd et al. (2011); Dean et al. (2014); Hill et al. (2009)). Despite a large theoretical literature proposing a link between individual cognitive abilities and population-level cumulative culture, (e.g., Boyd & Richerson, (1996); Tennie et al. (2009); Tomasello et al. (2005)), there is relatively little supporting experimental evidence, both in terms of what mechanisms support cumulative culture, how, and under what circumstances.

The apparent distinctiveness of mental state understanding to humans was the main motivator in exploring this capacity in relation to cumulative culture. Mental state understanding is not only distinct in humans, but also develops relatively late in human ontogeny (discussed fully in Chapter 1), further supporting the view that it may be rare or absent in other species. Mental state understanding also has multiple possible benefits for cumulative culture, all of which were explored to some extent in this thesis:

Social learning

Given the lack of evidence for cumulative culture in non-humans, it is unlikely that social learning strategies applied in non-human populations could act as a facilitator. Non-human (and a lot of human) social learning is driven by heuristic biases (Heyes & Pearce, 2015), however there may be some cases in which human social learning is driven by a reasoned understanding of what the social model knows (i.e., mental state understanding). Social learning driven by reasoned understanding is likely to be far more effective in the context of cumulative culture compared with biases that are driven by associative learning and/or biological adaptions.

While this thesis did not directly explore the benefit of reasoned understanding over associative learning, the idea that humans do adopt a reasoned strategy for learning from others ran concurrent through all experimental chapters in this thesis.

Flexible teaching

Teaching, as defined from a functional perspective, is rare in non-humans, and where it does occur (e.g., Franks & Richardson (2006); Thornton & MaCuliffe (2006)), it is largely restricted to a particular context. This means that the teacher is only able to teach a specific skill in a specific scenario, whereas human teaching spans a wide range of contexts. Such generalizable teaching is argued to require assumptions to be made about the knowledge of others, which allows the teacher to tailor signals to the needs of the learner (Thornton & Raihani, 2008). The general-purpose nature of human teaching likely arises as a consequence of an individual's recognition that their actions can potentially benefit others' learning, along with some understanding of how this might be achieved. In contrast it is widely agreed that non-human teaching doesn't involve this kind of understanding, due to disparities in capacities for understanding mental states (e.g., Kline (2015); Thornton & Raihani (2008)).

Previous research capturing the benefits of human teaching focuses broadly on whether teaching itself is more useful for information transmission than other social mechanisms, and these teaching conditions often involve sharing more information than is available to the learner in other conditions (Caldwell & Millen, 2009; Zwirner &

Thornton, 2015). As such, these studies do not highlight, specifically, the teacher's understanding of the learner's needs. Therefore, this thesis aimed to capture realistic teaching scenarios, in which a teacher can only transmit a limited amount of information to a learner, thus allowing an assessment of the choices teachers made about what to transmit. Furthermore, this thesis aimed to highlight the benefit of intentionally taught information by comparing this condition with an equivalently sized transmission of inadvertent information sharing. Even without restrictions on time, it would be impossible for a cultural parent to transmit every detail of the knowledge they have to the cultural offspring. Therefore, it holds more benefit for these transmissions to be made in a way that inspires inference of additional information from the surface transmission, in order to support successful replication and accumulation. This inference is likely to be built upon mental state understanding, as it requires the cultural parent to have an understanding of the cultural offspring's knowledge state, in order for them to select specific information that can achieve this.

This thesis aimed to capture the benefits of intentional knowledge sharing, both in terms of supporting cultural offspring to repeat beneficial behaviours, and to accumulate beneficial information over generations, reminiscent of cumulative culture.

Bidirectional communication

The use of intentional knowledge sharing for supporting cumulative culture is not exclusive to the cultural parent. It may also be necessary for the cultural offspring to select information about their own knowledge to guide the transmissions made by the cultural parent. While there is very little experimental evidence supporting this claim (as discussed in Chapters 1 and 5), it is widely accepted, and supported in Chapters 3 and 4, that humans can access mental state understanding to produce tailored signals. However, it may be possible that passively produced signals from cultural offspring are just as beneficial for conveying their knowledge state to a cultural parent.

Therefore, this thesis aimed to assess whether intentional knowledge sharing, relative to inadvertent knowledge sharing from cultural offspring facilitated effective knowledge sharing from the cultural parent, and ultimately whether this exchange supported cumulative culture.

Summary of the main questions this thesis aimed to answer

Overarching Question:

Does mental state understanding facilitate distinctively human cumulative culture?

Contributing questions addressed in individual chapters:

1. Proof of principle: Are human adults able to utilise information from a cultural parent in such a way that they can outperform a model demonstration at various levels of difficulty?

    1.1. Can we capture cumulative culture by using a selection task involving high-precision vicarious information?

    1.2. Do participant's performance increases amount to accumulation over simulated generations of learners?

2. Does intentional knowledge sharing from a cultural parent to a cultural offspring allow accumulation of knowledge over and above instances of more inadvertent knowledge sharing (representative of non-human teaching)?

3. Do the benefits of intentional knowledge sharing from a cultural parent to a cultural offspring emerge over a developmental trajectory that corresponds to advances in mental state understanding?

4. Does intentional knowledge sharing from the cultural offspring to the cultural parent further facilitate the cultural parent's ability to select meaningful and tailored information to support the cultural offspring's learning?

The following sections contain summaries of the aims, methods, results and contributions of all four of the experimental chapters in this thesis. These summaries are grouped by methodology used, limitations of specific methodologies, and future directions of study are discussed at the end of each section.

6.2. Demonstrating a new methodology for examining performance improvement over generations of social transmission…without social transmission

6.2.1 Summary of Chapter 2

In Chapter 2, we conducted a proof of principle to demonstrate that human adults can utilise information from a cultural parent to accumulate information over generations, and that this effect can be captured using a selection task feeding participants vicarious information from a computer-generated cultural parent model (‘*parent model*’).

The other question in this study was whether it is possible to use individual-level data to infer the likely outcome of a transmission chain. To do this, participants were exposed to vicarious information equivalent to a range of different levels of success, which could then be modelled to generate multi-generational transmission.

This was investigated this in two realistic contexts whereby the vicarious information was either of high- or low-precision. In the High-precision condition, more information about the parent model search was available in the vicarious information, specifically whether their individual selections were rewarded. The Low-precision condition did not carry this benefit, and participants were only afforded the location of the parent model search and the outcome of the entire grid search (i.e., no information about the value of specific search locations). These settings of various precision corresponded to different ways in which cumulative culture may operate in humans, i.e., probably sometimes caused by insightful and deliberate accumulation (High-precision condition) (Osiurak et al., 2016) but perhaps in other situations arising from the accumulation of adaptive traits as a result of biases in high-fidelity transmission, accompanied by errors and random innovations (Low-precision condition) (Derex, Bonnefon, Boyd, & Mesoudi, 2019).

This experiment used a simple computer-based task in which participants were asked to search reward landscapes of multiple sizes for rewards, using vicarious information from a parent model to guide them. All possible reward outcomes for the parent model were shown to participants, meaning that they were effectively creating a response to vicarious information at all generations. To assess whether performance increase at the individual level amounted to cumulative culture at the population level, we simulated transmission chains using the participant's behavioural data. Therefore, along with testing for the human propensity to see value in information provided by others, this task had the additional benefit of testing the novel '*potential for ratcheting'* paradigm outlined by Caldwell et al. (2020).

In the High-precision condition, we found that generally, adult participants largely repeated rewarded parent model selections, and avoided unrewarded parent model selections in favour of individual exploration. As a result of adopting this insightful strategy, we found that participants either matched or outperformed the vicarious information from the parent model in almost every landscape size. Consistent with this, when the behavioural data was used to simulate multi-generational transmission, population chains accumulated at a similar rate to chains built to strictly adopt the low-risk behavioural strategy in almost every landscape size.

In the Low-precision condition, there was much less scope for individual participants to adopt a rule that ensured accumulation of beneficial information in most cases. The only rule that would guarantee no loss of information from the parent model was to copy all selections if all, or even *some,* rewarded selections had been made by the parent model. Similarly, the only situation in which participants should have risked deviating from the exact selections made by the parent model was in circumstances where the parent model scored below chance, and so scoring lower was unlikely with an avoid-all strategy. This strategy did, however, restrict grid maximisation to limited circumstances where avoiding all of the parent model selections resulted in selection of all targets by chance, which could only happen when the parent model's score was zero, and was in all cases relatively unlikely, especially for some grid sizes. In some cases, particularly in circumstances where only some of the parent selections were rewarded, real participant behaviour in fact deviated from this low-risk strategy, and involved a '*mixed strategy*', whereby some parent selections were copied, and others were avoided. Of course, because there was no indication as to which selections were rewarded, the strategy was applied somewhat arbitrarily, and on average this strategy could never outperform the low-risk repeat-all/avoid-all strategy as perfectly applied (which may not have been possible for real participants anyway, as it involved calculating the exact tipping point). Perhaps somewhat counterintuitively however, this mixed strategy meant that result of the transmission chain simulation from the behavioural data had the potential to reach the maximum score faster than the low-risk strategy, because, although individual participants on average performed more poorly than the low-risk strategy, they were also increasing their opportunities to select all of the rewarded stimuli by chance. Since even in real participant behaviour, a perfect parent model score usually resulted in a repeat-all response, these maximum scores – once achieved – were then relatively well-preserved within the chain simulations. Nonetheless, the likelihood of maximizing the score of the grid remained relatively improbable, and increasingly so as grid sizes increased.

6.2.2 Chapter 2: Contributions to current literature

The methodology used in this chapter demonstrates an important new principle that could influence the field at a broad domain. As discussed in Chapter 2, traditional methods for capturing cumulative culture in lab studies are very restrictive for use with adult populations due to practical limitations, which affect use with other populations (children and non-humans) further. By using behavioural data collected from every possible transmission outcome, chains of transmission can easily be simulated without the need for the number of participants and complex organisation that goes along with necessarily extensive requirements of traditional tasks involving real transmission between different individuals. The current methodology could be adapted for use in future research, to further explore the transmission conditions under which cumulative culture can occur in human adults, children and non-humans.

The findings of this chapter served to inform the later chapters in this thesis by confirming that human adults can accumulate information over simulated multi-generational transmission chains in a computer-based task, making few errors when adopting a reasoned strategy. However, we did not adopt the specific PFR simulation method for any of the following tasks. The general task approach was largely the same, and used a grid search task with high-precision vicarious information. However, in later tasks, we increased the complexity of the problem space, such that targets appeared in clusters and could be located anywhere on the grid, rather than one per row. As such, many factors, such as the shape of targets and their position relative to one another, could influence the cultural parent's decision about what to transmit. Generating all possible scenarios of vicarious information and having participants respond to every one would be much more difficult than gleaning this information from a traditional transmission chain design. Therefore, while there is an inherent benefit to using the '*potential for ratcheting'* paradigm in terms of lowering participant recruitment and reducing organisation (as discussed in Chapter 2), we deemed it too important to have vicarious information come from a participating cultural parent. This is because, in the current study, we aimed to demonstrate accumulations, however in later studies we aimed to assess transmission behaviours, including biases that would be difficult to replicate in a computer-generated cultural parent model.

6.3 Intentional information sharing promotes cumulative culture relative to inadvertent behavioural cues

6.3.1 Summary of Chapter 3

In Chapter 2, it was established that adult participants were able to significantly outperform a cultural parent such that cumulative culture can be observed in simulated multi-generational transmission chains. Chapter 3 used a grid search task to directly compare conditions in which information *transmitted* from human cultural parent to cultural offspring over multiple transmission generations affected the population-wide ability to accumulate information over those generations. Specifically, we aimed to establish, in conditions where the amount of transmission is limited, whether intentional knowledge sharing from a cultural parent to a cultural offspring allows accumulation of knowledge over and above instances of more inadvertent knowledge sharing (representative of non-human teaching).

The experiment used in this task involved transmission of grid search information from cultural parents to cultural offspring in transmission chains made up of ten participants. The information transmitted from cultural parents was manipulated such that cultural offspring received a limited subset of intentionally produced information (Intentional condition), a limited subset of information randomly generated from the cultural parent's search (Inadvertent condition), or the cultural parent's full search attempt (Full condition). The type of information transmission (Intentional, Inadvertent or Full) was consistent across chains, therefore we ran five chains in each of the three conditions. All information was of high precision, therefore the value of uncovered grid squares in the transmission was made available to cultural offspring.

The results of this study revealed that a limited subset of intentional knowledge transmission facilitated accumulation of knowledge over generations at a similar rate as when full information from the cultural parent's search was made accessible to the cultural offspring. Comparatively, inadvertent knowledge transmission produced far more limited accumulation over generations. We concluded that in cases where only a small amount of information can be transmitted, there is an inherent benefit to cumulative culture if that information was intentionally produced by the cultural parent. In further exploratory analysis, we attributed the success of Intentional chains to the cultural parent's transmission preferences, as they appeared to have made effective choices about what specific information would benefit the cultural offspring both in terms of information accumulation and reduction of loss.

6.3.2 Chapter 3: Contributions to current literature

This chapter added to the relatively young literature using lab-based transmission studies to manipulate learning conditions with the aim to explore which of those conditions are the most suitable for supporting cumulative culture. A recent study by Lucas et al. (2020) highlighted the benefits of teaching in causally opaque tasks, with transmission chains in which cultural parents instructed the cultural offspring accumulating information much faster than conditions of individual learning. However, this study, and its notable predecessors (Caldwell & Millen, 2009; Zwirner & Thornton, 2015) focused primarily on teaching as a general benefit, rather than the specific actions of the teacher that produce this benefit, which meant that teaching conditions often resulted in a higher volume of information being transmitted between participants. Therefore, we extended this literature by controlling for the volume of transmitted information, which allowed a direct comparison of intentional teaching versus inadvertent transmissions.

Our findings suggested that there are some conditions in which intentional knowledge sharing facilitates cumulative cultural evolution, however we still weren't certain that the apparently strategic choices made by cultural parents could have occurred due to a sophisticated understanding of the cultural offspring's mental state, or for some other reason (e.g., inherent preference for selecting central tiles, and for selecting hits over misses, which could have still facilitated learning in spite of not being designed for that purpose).

6.3.3 Summary of Chapter 4

In Chapter 3, adults' ability, as a cultural parent, to transmit limited information that facilitated the performance of a cultural offspring was attributed to their complex understanding of the cultural offspring's knowledge state. As such, cultural parents selected information for transmission that would be particularly beneficial to a naïve cultural offspring; however, the role of advanced cognition in this task was largely speculative. Chapter 4 aimed to extend this finding, particularly to investigate the age at which this beneficial information transmission began to emerge. The results of this task in a child sample can therefore be used as a means of determining whether this effect might be linked to cognitive capacities that develop in humans at a young age, such as advances in mental state understanding. Such a finding would bolster the theory that this is distinctively human.

In the existing literature, children's teaching behaviour is generally found to show age-related changes in line with the emergence of mental state understanding (Ronfard &

Corriveau, 2016; Wood et al., 1995; Ziv et al., 2016). As such, we decided that extending the methodology used in Chapter 3 to a sample of children would provide an important insight into whether this method could capture an age-related increase in cumulative culture. This chapter therefore aimed to support the idea that cumulative culture is distinctively human and *relies* on capacities that develop in early childhood.

We used a paradigm adapted from the one that was used in Chapter 3 to test a large sample of children aged 5 to 10 years old. Minor methodological edits were made to minimise the difficulties associated with testing children of multiple age groups.

We found a very clear developmental trajectory for the emergence of the transmission preferences, and resulting accumulation, observed in the adults. Chains of younger children produced no accumulation in any condition, whereas chains of older children produced a pattern of results that mirrored that of the adults in Chapter 3. This result supported the speculative interpretation of the data from Chapter 3, that the ability to exhibit distinctively human cumulative culture is likely to be connected with the emergence of mental state understanding.

Additionally, while all age groups had a significant bias for transmitting information about target grid squares compared with non-target grid squares, only children in the older age categories had a significant preference for transmitting the central target segment over any other target segment, and this preference increased further between the 7- to 8-year-olds and the 9- to 10-year-olds. This supports the theory that specific selections made by adults were driven by some insight into their value to a potential learner. Cultural parents made selections specifically tailored to the needs of the cultural offspring, and it is likely that they did this using a sophisticated understanding of the cultural offspring's knowledge state.

6.3.4 Chapter 4: Contributions to current literature

In the current literature, there is a very limited pool of transmission chain studies that involve child participants, and those that do (e.g., Reindl & Tennie (2018), chains of 4- to 6-year-olds) find no evidence that children can improve (additively) on an earlier generation's example using chain designs. The study in this chapter not only successfully identified cumulative culture in children, but also found a clear developmental trajectory of the capacity to utilise intentional information from a cultural parent to increase cumulative culture. This suggests that children are not able to accumulate information over generations prior to the development of cognitive capacities known to emerge at around 4 years of age. Given that children were only able to accumulate information from full cultural parent performance at 7 years of age, this suggests that these capacities have to be somewhat established before they become effective enough to facilitate cumulative culture.

6.3.5 Concluding summary of Chapters 3 and 4

Together, the studies conducted in Chapters 3 and 4 lend great support to the current literature of both mental state understanding and cumulative culture. We not only established a methodology to assess conditions under which cumulative culture can occur, but used this to capture circumstances in which restricting the use of distinctively human cognitive capacities at an individual level reduces the ability to accumulate information over generations at the population level. We bolstered this finding by capturing a developmental trajectory whereby downward transmissions of full cultural parent performance facilitated cumulative culture at 7 to 8 years of age. However, adult-like accumulation, whereby transmission of a small subset of intentionally produced information is able to facilitate cumulative culture as efficiently as full information, only emerged in chains of 9- to 10-year-olds. While these results support the role of distinctively human mechanisms in cumulative culture, they do not specifically isolate mental state understanding, and so to conclude that a fully developed and sophisticated mental state understanding facilitates cumulative cultural evolution based on a correlation with age, would be very speculative. In Chapter 4, we did attempt to employ a test of mental state understanding with children, with the hopes of supporting the role of mental state understanding further. However, due to the fragility of the task itself, no clear conclusions could be drawn. Of course it is entirely possible that the developmental trajectory found in Chapter 4 could potentially have been caused by any of the suite of cognitive mechanisms that emerge at around the same age. In particular, higher orders of mental state understanding and perspective taking, both of which are found to develop at around the same age as we begin to see adult-like performance in the child task.

6.4 The effect of cultural offspring feedback on cumulative cultural evolution in a grid search task

6.4.1 Summary of Chapter 5

Because most instances of human teaching involve bidirectional transfer of information between cultural offspring and cultural parent (such that the offspring provides feedback or asks questions), Chapter 5 extended the results of the previous studies in this thesis by exploring the importance of bidirectional knowledge transfer in cumulative culture.

In this study, we preserved only the intentional knowledge transmission condition from the cultural parent, because based on the results of Chapter 3, we had established that this condition facilitated cumulative culture. Furthermore, manipulating the cultural parent's transmissions using any of the other conditions used previously would not allow an assessment of whether feedback from the cultural offspring guides transmissions made by the cultural parent. Therefore instead, we manipulated the intentionality of upward transmissions – that is, transmissions made from the cultural offspring to the cultural parent. We used the similar conditions as in the two preceding chapters such that cultural offspring information either consisted of the cultural offspring's full search, or a subset of three of the cultural offspring's search either selected by the offspring or randomly generated. We also included an additional condition in which no cultural offspring information was transmitted. This was to provide a low-bar baseline on which to compare the other conditions. We hypothesised that conditions in which full information, or an intentional subset of information was transmitted from cultural offspring to cultural parent would facilitate cumulative culture over and above what was possible in unidirectional, or inadvertent feedback, transmission conditions. This chapter therefore aimed to establish whether cultural offspring can make effective strategic decisions about what to transmit to the parent about their own knowledge, to in turn inform the parent about what information they require.

The methodology for this task was similar to the previous tasks such that they required participants to search grids to reveal targets. Selection of target segments amounted to an overall score per searched grid, and the increase of this score over generations of learners was measured to explore cumulative culture. However, in this task, rather than cultural parents giving information to cultural offspring in a single unidirectional transmission, three separate (downward) parent transmissions occurred between (upward) cultural offspring transmissions, such that cultural parent transmissions were made in *response* to cultural offspring transmissions. The final score for cultural offspring grids was collected after their final search, which followed the final transmission from the cultural parent.

The results of this chapter revealed no effect of offspring feedback on cumulative culture, with chains in all four conditions showing an equal accumulation over generations. Because one of these conditions delivered no feedback from the cultural offspring to the cultural parent, we attributed this accumulation to the intentionality of the information transmitted from the cultural parent alone, and concluded that the design was in fact not sufficiently sensitive to detect an effect of any differences between the intentional and inadvertent feedback conditions.

6.4.2 Design limitations in Chapter 5, and suggestions of future study to mediate these

There are several possible reasons why we didn't find any differences in score increase between conditions in this task. The first is that, very simply, cultural offspring feedback has absolutely no effect, and that accumulation in this task rests solely on intentional information being transmitted from the cultural parent to the cultural offspring. It could be possible that there was simply not enough motivation for offspring to tailor their transmissions in a way that would help them gain more information. As such, in the discussion of Chapter 5, we made some suggestions for how to adapt this study to ensure all information send between dyads held some value.

It is more likely, however, that an effect was not able to be seen because of task design. This was made prominent by the lack of difference between the Full and Null transmission conditions, which were expected to reflect opposite extremes, even if it was attributable to information volume rather than intentionality. It was suspected that this happened because in a lot of cases, the cultural parent's potential to impart beneficial knowledge had reached its maximum well before their final opportunity for communication (i.e., through transmitting or receiving information about all of the ships they know about).

In future research, it might be possible to rectify this by altering the parameters of the task to reduce the possibility of cultural offspring finding all of the rewards discovered by their cultural parent before the cultural parent had completed their downward transmissions.

6.5 Benefits of the methodological approach used

The methodology that we used throughout this thesis opens up possibilities for future research in a broad scope. Grid search tasks are beneficial because they can be tightly controlled with respect to the information sent between participants. They also allow systematic analysis of search biases and transmission biases, and can be applied to many different experimental scenarios to explore transmission conditions for cumulative culture. The task itself is also very malleable, for example, in the discussion of Chapter 5, we proposed a version of the grid search task that would invert the logic of the grid search task, such that instead of searching for targets, the object was to avoid obstacles, with participants goal being to move one grid square at a time from one end of the grid to the other. Such changes to the task design could greatly broaden its potential for testing novel hypotheses.

There are also a lot of practical benefits of using grid search tasks. In Chapter 2, we discussed practical limitations of traditional transmission chain designs, which included recruitment and organisation of participants such that they are able to see the process or outcome of the previous participant's performance to scaffold their own. By using computer-based grid search tasks, it's possible to have multiple chains and conditions running at once, because all search data and transmission data can be stored efficiently and automatically. For example, in the experimental task used in Chapter 3, we had three chains running at any given time, in order to pseudo-randomly assign participants to generations. In Chapter 4, we had the additional restriction of age category, however, given that there was no limit to the number of chains run consecutively at any given time, this was unproblematic. There were slightly more restrictions introduced in Chapter 5 given that adjacent members of the same chain had to take part simultaneously, however, we were still able to run multiple chains, and so, assess multiple transmission conditions at the same time – culminating in the total testing time for 243 participants coming out as just under five days.

Running the task on touch-screen tablets holds the additional benefit of portability. Testing sessions can be run using touch screen laptops or tablets in many locations. For example, we collected data for Chapter 3 in the University of Stirling and at Glasgow science centre, and the data for Chapter 4 was collected from six different Highland schools. The data for Chapter 4 is an excellent example of multiple chains running at once, because all age groups and conditions were tested in each school – meaning at least nine chains were running *between schools*. While this task design doesn't necessarily get around the issue of large participant pools, it does eliminate any need for complex organisation.

There may be some circumstances in which collecting grid search data does hold a benefit for reducing participant numbers. In Chapter 2, we assessed a new methodology which used model participant data to create large datasets of behavioural responses which could be arranged into chains to assess the potential for ratcheting over generations. It is entirely possible that, in some contexts, the data collected in this study and future studies like it could be used to assess additional constraints or biases on the selection of information for transmission.

6.6 New contributions to the literature that support this research

There is extensive research focus in mental state understanding, and while this thesis gives a current account of the relevant research available at the time, the field is in a state of constant development. Research in the field of cumulative culture is relatively young, however, the field is getting closer to unravelling the questions around whether it is distinctively human, and if so, what capacities humans have that give access to it, or restrict it in non-humans. The current thesis adds both to the growing support for mental state understanding as a distinctively human mechanism, and as a facilitator of uniquely human cumulative culture.

Since this thesis has been in production, little has changed in the fields of either mental state understanding or cumulative culture that would affect the methodology I used, or the motivations under which I ran the studies. However there have been some recent theoretical advancements that give added weight to the ideas this work was based on. Part of the argument for mental state understanding as a distinctively human mechanism comes from its late ontogenetic development, however there are a number of studies that have claimed to find evidence of mental state understanding using indirect behavioural metrics, rather than direct explicit questions (see Chapter 1 for a full summary of these). Although non-mentalistic alternative explanations for these results span back as far as the original implementation of the studies, over the last few years, a number of replications have been attempted, many of which have failed to capture the same results using an identical paradigm (Kulke & Rakoczy, 2018). These failures cast further doubt on these findings, and thus give greater credence to the view that mental state understanding is a late-developing system. This reinforces the likelihood that it could be a distinctively human capacity, in line with the rational for the work presented in the current thesis.

6.7 General conclusions

The main goal of this thesis was to contribute to the existing literature exploring distinctively human cognitive mechanisms. Specifically, the aim was to investigate whether and how the distinctively human propensity for understanding the mental states of others facilitates distinctively human cumulative cultural evolution. Four experimental studies were conducted, each employing a novel grid search task to assess the benefits for cumulative culture of intentionally produced, over inadvertently produced information from a cultural parent. Overall, the studies presented in this thesis suggest that intentional downward information transmission can support cumulative culture relative to circumstances where only inadvertent information is available to a learner. This thesis also supported the idea that the capacity to transmit, and potentially make use of, this intentional information increases with age in a similar trajectory to higher-order mental state understanding. However, little support was found for a further facilitatory effect of cultural offspring feedback. The results of these studies together lend support to the theory that cumulative culture is facilitated by higher order mental state understanding; which in turn, supports the theory that cumulative cultural evolution is unique to humans.

References

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953.

Ashley, J., & Tomasello, M. (1998). Cooperative Problem-Solving and Teaching in Preschoolers. *Social Development*, *7*(2), 143–163. https://doi.org/10.1111/1467-9507.00059

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, *35*(5), 1311–1320. https://doi.org/10.1037/0012-1649.35.5.1311

Atkinson, M., Blakey, K. H., & Caldwell, C. A. (2020). Inferring Behavior From Partial Social Information Plays Little or No Role in the Cultural Transmission of Adaptive Traits. *Cognitive Science*, *44*(10). https://doi.org/10.1111/cogs.12903

Baron-cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, *21*(1), 37–46.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Basalla, G. (1988). *The Evolution of Technology*. Cambridge University Press.

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

Baum, W. M., Richerson, P. J., Efferson, C. M., & Paciotti, B. M. (2004). Cultural evolution in laboratory microsocieties including traditions of rule giving and rule following. *Evolution and Human Behavior*, *25*(5), 305–326. https://doi.org/10.1016/j.evolhumbehav.2004.05.003

Bebbington, K., MacLeod, C., Ellison, T. M., & Fay, N. (2017). The sky is falling: evidence of a negativity bias in the social transmission of information. *Evolution and Human Behavior*, *38*(1), 92–101. https://doi.org/10.1016/j.evolhumbehav.2016.07.004

Blakey, K. H., Renner, E., Atkinson, M., Rafetseder, E., & Caldwell, C. A. (Under Review). *Age-related transition from simple associations to reasoning-based social learning strategies in 4- to 8-year-old children (Homo sapiens).*

Boogert, N. J., Reader, S. M., Hoppitt, W., & Laland, K. N. (2008). The origin and spread of innovations in starlings. *Animal Behaviour*, *75*(4), 1509–1518. https://doi.org/10.1016/j.anbehav.2007.09.033

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

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

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning Memory and Cognition*, *22*(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Bridgers, S., Jara-Ettinger, J., & Gweon, H. (2016). Children Consider Learners' Costs and Rewards When Deciding What to Teach. *CogSci*.

Buckner, C. (2014). The Semantic Problem(s) with Research on Animal Mind-Reading. *Mind & Language*, *29*(5), 566–589.

Burdett, E. R. R., Dean, L. G., & Ronfard, S. (2018). A Diverse and Flexible Teaching Toolkit Facilitates the Human Capacity for Cumulative Culture. *Review of Philosophy and Psychology*, *9*(4), 807–818. https://doi.org/10.1007/s13164-017-0345-4

Burnside, K., Ruel, A., Azar, N., & Poulin-dubois, D. (2018). Implicit false belief across the lifespan : Non-replication of an anticipatory looking task. *Cognitive Development*, *46*(February 2017), 4–11. https://doi.org/10.1016/j.cogdev.2017.08.006

Buttelmann, D., Buttelmann, F., Carpenter, M., Call, J., & Tomasello, M. (2017). Great apes distinguish true from false beliefs in an interactive helping task. *Plos One*, *12*(4), e0173793. https://doi.org/10.1371/journal.pone.0173793

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*(2), 337–342. https://doi.org/10.1016/j.cognition.2009.05.006

Byrne, R. W., & Whiten, A. (1992). Cognitive Evolution in Primates : Evidence from Tactical Deception. *Man*, 609–627.

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), 1–17. https://doi.org/10.1111/tops.12391

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

Caldwell, C. A., & Millen, A. E. (2008a). 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. (2008b). *Studying cumulative cultural evolution in the laboratory*. *September*, 3529–3539. https://doi.org/10.1098/rstb.2008.0133

Caldwell, C. A., & Millen, A. E. (2009). Social Learning Mechanisms and Cumulative Cultural Evolution. *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*, *9*(4), 751–770. https://doi.org/10.1007/s13164-017-0346-3

Call, J. (2001). Body imitation in an enculturated orangutan (pongo pygmaeus). *Cybernetics and Systems*, *32*(1–2), 97–119.

https://doi.org/10.1080/019697201300001821

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

Call, J., & Tomasello, M. (1999). *A Nonverbal False Belief Task : The Performance of Children and Great Apes*. *70*(2), 381–395.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*(4), 1032–1053. https://doi.org/10.1111/1467-8624.00333

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development: An International Journal of Research and Practice*, *11*(2), 73–92. https://doi.org/10.1002/icd

Caro, T. M., & Hauser, M. D. (1992). Is There Teaching in Non-human Animals? *The Quarterly Review of Biology*, *67*(2), 151–174. https://doi.org/10.1086/417553

Claidière, N., Amedon, G. K., André, J.-B., Kirby, S., Smith, K., Sperber, D., & Fagot, J. (2018). Convergent transformation and selection in cultural evolution. *Evolution and Human Behavior*, *39*(2), 191–202. https://doi.org/10.1016/j.evolhumbehav.2017.12.007

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), 1–9. https://doi.org/10.1098/rspb.2014.1541

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*(4), 377–395. https://doi.org/10.1016/0885-2014(94)90012-4

Cole, G. G., Atkinson, M., Le, A. T. D., & Smith, D. T. (2016). Do humans spontaneously take the perspective of others? *Acta Psychologica*, *164*, 165–168. https://doi.org/10.1016/j.actpsy.2016.01.007

Conway, J. R., Lee, D., Ojaghi, M., Catmur, C., & Bird, G. (2017). Submentalizing or mentalizing in a level 1 perspective-taking task: A cloak and goggles test. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(3), 454–465. https://doi.org/10.1037/xhp0000319

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

Custance, D. M., Whiten, A., & Bard, K. A. (1995). Can young chimpanzees (Pan troglodytes) imitate arbitrary actions? Hayes & Hayes (1952) revisited. *Behaviour*, *132*, 11–12. http://www.jstor.org/stable/10.2307/4535304

Danchin, É., Giraldeau, L. A., Valone, T. J., & Wagner, R. H. (2004). Public information: From nosy neighbors to cultural evolution. *Science*, *305*(5683), 487–491. https://doi.org/10.1126/science.1098254

Davis-Unger, A., & Carlson, S. M. (2008). Development of teaching skills and relations to theory of mind in preschoolers. *Journal of Cognition and Development*, *9*(1), 26–45. https://doi.org/10.1080/15248370701836584

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1-12. doi:10.3758/s13428-014-0458-y.

De Villiers, J. (2007). The interface of language and Theory of Mind. *Lingua*, *117*(11), 1858–1878. https://doi.org/10.1016/j.lingua.2006.11.006

De Villiers, J., & De Villiers, P. A. (2014). The role of language in theory of mind development. *Topics in Language Disorders*, *34*(4), 313–328. https://doi.org/10.1097/TLD.0000000000000037

Dean, L. G., Hoppitt, W., Laland, K. N., & Kendal, R. L. (2011). Sex ratio affects sex-specific innovation and learning in captive ruffed lemurs (Varecia variegata and Varecia rubra). *American Journal of Primatology*, *73*(12), 1210–1221.

https://doi.org/10.1002/ajp.20991

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*, *178*(1974), 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

Dennett, D. (1978). Beliefs about Beliefs. *Behavioral and Brain Sciences*, *1*, 568–570.

Derex, M., Bonnefon, J. F., Boyd, R., & Mesoudi, A. (2019). Causal understanding is not necessary for the improvement of culturally evolving technology. *Nature Human Behaviour*, *3*(5), 446–452. https://doi.org/10.1038/s41562-019-0567-9

Devine, R. T., & Hughes, C. (2016). Measuring theory of mind across middle childhood: Reliability and validity of the Silent Films and Strange Stories tasks. *Journal of Experimental Child Psychology*, *149*, 23–40. https://doi.org/10.1016/j.jecp.2015.07.011

Doherty, M. J. (2008). *Theory of mind: How children understand others' thoughts and feelings.* Psychology Press.

Doherty, M. J., & Wimmer, M. C. (2005). Children's understanding of ambiguous figures: Which cognitive developments are necessary to experience reversal? *Cognitive Development*, *20*(3), 407–421. https://doi.org/10.1016/j.cogdev.2005.05.003

Dörrenberg, S., Rakoczy, H., & Liszkowski, U. (2018). How (not) to measure infant Theory of Mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development*, *46*(February), 12–30. https://doi.org/10.1016/j.cogdev.2018.01.001

Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, *13*(2), 331–338. https://doi.org/10.1111/j.1467-7687.2009.00888.x

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

Fischer, A. L., O'Rourke, N., & Thornton, W. L. (2017). Age differences in cognitive and affective theory of mind: Concurrent contributions of neurocognitive performance, sex, and pulse pressure. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *72*(1), 71–81. https://doi.org/10.1093/geronb/gbw088

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1-Level 2 distinction. *Developmental Psychology*, *17*(1), 99–103. https://doi.org/10.1037/0012-1649.17.1.99

Fogarty, L., Strimling, P., & Laland, K. N. (2011). The evolution of teaching. *Evolution*, *65*(10), 2760–2770. https://doi.org/10.1111/j.1558-5646.2011.01370.x

Franks, N. R., & Richardson, T. (2006). Teaching in tandem-running ants. *Nature*, *439*(7073), 153. https://doi.org/10.1038/439153a

Furlanetto, T., Becchio, C., Samson, D., & Apperly, I. A. (2016). Altercentric interference in Level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. *Journal of Experimental Psychology : Human Perception and Performance*, *42*(2), 158. https://doi.org/http://dx.doi.org/10.1037/xhp0000138

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

Garland, E. C., Goldizen, A. W., Rekdahl, M. L., Constantine, R., Garrigue, C., Hauser, N. D., Poole, M. M., Robbins, J., & Noad, M. J. (2011). Dynamic horizontal cultural transmission of humpback whale song at the ocean basin scale. *Current Biology*, *21*(8), 687–691. https://doi.org/10.1016/j.cub.2011.03.019

Garnham, W. A., & Ruffman, T. (2001). Doesn't see, doesn't know: Is anticipatory looking really related to understanding of belief? *Developmental Science*, *4*(1), 94–100. https://doi.org/10.1111/1467-7687.00153

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181–218.

Gergely, G., & Csibra, G. (1997). Teleological reasoning in infancy: The infant's naive theory of rational action a reply to Premack and Premack. *Cognition*, *63*(2), 227–233. https://doi.org/10.1016/S0010-0277(97)00004-8

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292. https://doi.org/10.1016/S1364-6613(03)00128-1

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165–193. https://doi.org/10.1016/0010-0277(95)00661-H

Gomez, J. C. (2004). *Apes, monkeys, children, and the growth of the mind.* Harvard Universoty Press.

Gopnik A., Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, *59*(1), 26–37. https://doi.org/10.1111/j.1467-8624.1988.tb03192.x

Gruber, T. (2016). Great Apes Do Not Learn Novel Tool Use Easily: Conservatism, Functional Fixedness, or Cultural Influence? *International Journal of Primatology*, *37*(2), 296–316. https://doi.org/10.1007/s10764-016-9902-4

Gweon, H., Chu, V., & Schulz, L. E. (2014). To give a fish or to teach how to fish? Children weigh costs and benefits in considering what information to transmit. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*(36).

Gweon, H., Shafto, P., & Schulz, L. E. (2014). Children consider prior knowledge and the cost of information both in learning from and teaching others. *Proceedings of the Annual Meeting of the Cognitive Science Behaviour*, *36*(36), 565–570. https://doi.org/https://doi.org/ISBN 978-0-9768318-8-4

Hall, K., Oram, M. W., Campbell, M. W., Eppley, T. M., Byrne, R. W., & de Waal, F. B. M.

(2017). Chimpanzee uses manipulative gaze cues to conceal and reveal information to foraging competitor. *American Journal of Primatology*, *79*(3), 1–11. https://doi.org/10.1002/ajp.22622

Happé, F. G. E. (1994). An Advanced Test of Theory of Mind: Understanding of Story Characters' Thoughts and Feelings by Able Autistic, Mentally Handicapped, and Normal Children adn Adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129–254. https://doi.org/10.1016/S0214-4603(10)70120-0

Hayashi, T., Akikawa, R., Kawasaki, K., Egawa, J., Minamimoto, T., Kobayashi, K., Kato, S., Hori, Y., Nagai, Y., Iijima, A., Someya, T., & Hasegawa, I. (2020). Macaques Exhibit Implicit Gaze Bias Anticipating Others' False-Belief-Driven Actions via Medial Prefrontal Cortex. *Cell Reports*, *30*(13), 4433-4444.e5. https://doi.org/10.1016/j.celrep.2020.03.013

Henrich, J., & Broesch, J. (2011). On the nature of cultural transmission networks: Evidence from Fijian villages for adaptive learning biases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1139–1148. https://doi.org/10.1098/rstb.2010.0323

Heyes, C. M. (1993). Imitation, culture and cognition. *Animal Behaviour*, *46*(5), 999–1010. https://doi.org/DOI 10.1006/anbe.1993.1281

Heyes, C. M. (1998). Theory of mind in non-human primates. *The Behavioral and Brain Sciences*, *21*(1), 101–148. http://www.ncbi.nlm.nih.gov/pubmed/10097012

Heyes, C. M. (2012). Simple minds: a qualified defence of associative learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2695–2703. https://doi.org/10.1098/rstb.2012.0217

Heyes, C. M. (2014). *Submentalizing : I Am Not Really Reading Your Mind*. https://doi.org/10.1177/1745691613518076

Heyes, C. M. (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. M., & Pearce, J. M. (2015). Not-so-social learning strategies. *Proceedings of* *the Royal Society of London B: Biological Sciences*, *282*(1802), 20141709. https://doi.org/10.1098/rspb.2014.1709

Heyes, C. M., Ray, E. D., Mitchell, C. J., & Nokes, T. (2000). Stimulus Enhancement: Controls for Social Facilitation and Local Enhancement. *Learning and Motivation*, *31*(2), 83–98. https://doi.org/10.1006/lmot.1999.1041

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

Hirata, S., & Matsuzawa, T. (2001). Tactics to obtain a hidden food item in chimpanzee pairs (Pan troglodytes). *Animal Cognition*, *4*(3–4), 285–295. https://doi.org/10.1007/s100710100096

Holzhaider, J. C., Hunt, G. R., & Gray, R. D. (2010). Social learning in New Caledonian crows. *Learning & Behavior*, *38*(3), 206–219. https://doi.org/10.3758/LB.38.3.206

Hopper, L. M., Flynn, E. G., Wood, L. A. N., & Whiten, A. (2010). Observational learning of tool use in children: Investigating cultural spread through diffusion chains and learning mechanisms through ghost displays. *Journal of Experimental Child Psychology*, *106*(1), 82–97. https://doi.org/10.1016/j.jecp.2009.12.001

Hopper, L. M., Schapiro, S. J., Lambeth, S. P., & Brosnan, S. F. (2011). Chimpanzees' socially maintained food preferences indicate both conservatism and conformity. *Animal Behaviour*, *81*(6), 1195–1202. https://doi.org/10.1016/j.anbehav.2011.03.002

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. J. E., Brown, G. R., Kendal, R., Rendell, L., Thornton, A., Webster, M. M., & Laland, K. N. (2008). Lessons from animal teaching. *Trends in Ecology and Evolution*, *23*(9), 486–493. https://doi.org/10.1016/j.tree.2008.05.008

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

Horschler, D. J., MacLean, E. L., & Santos, L. R. (2020). Do Non-human Primates Really Represent Others' Beliefs? *Trends in Cognitive Sciences*, *24*(8), 594–605. https://doi.org/10.1016/j.tics.2020.05.009

Hosey, G. R., Jacques, M., & Pitts, A. (1997). Drinking from tails: Social learning of a novel behaviour in a group of ring-tailed lemurs (Lemur catta). *Primates*, *38*(4), 415–422. https://doi.org/10.1007/BF02381881

Howe, N., Recchia, H., Porta, S. Della, & Funamoto, A. (2012). "The driver doesn't sit, he stands up like the Flintstones!": Sibling Teaching During Teacher-Directed and Self-Guided Tasks. *Journal of Cognition and Development*, *13*(2), 208–231. https://doi.org/10.1080/15248372.2011.577703

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

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

Jeong, J., & Frye, D. (2018). Explicit versus implicit understanding of teaching : Does knowing what teaching is help children to learn from it ? *Teaching and Teacher Education*, *71*, 355–365. https://doi.org/10.1016/j.tate.2018.02.002

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

Juvrud, J., & Gredebäck, G. (2020). The teleological stance: Past, present, and future. *Developmental Science*, e12970. https://doi.org/10.1111/desc.12970

Kano, F., Krupenye, C., Hirata, S., & Call, J. (2017). Eye tracking uncovered great apes' ability to anticipate that other individuals will act according to false beliefs. *Communicative and Integrative Biology*, *10*(2), 1–7. https://doi.org/10.1080/19420889.2017.1299836

Kano, F., Krupenye, C., Hirata, S., Tomonaga, M., & Call, J. (2019). Great apes use self-experience to anticipate an agent's action in a false-belief test. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(42), 20904–20909. https://doi.org/10.1073/pnas.1910095116

Karadeniz, P. G., & Ercan, I. (2017). Examining tests for comparing survival curves with right censored data. *Statistics in Transition*, *18*(2), 311–328. https://doi.org/10.21307/stattrans-2016-072

Kawai, M. (1965). Newly-acquired pre-cultural behavior of the natural troop of Japanese monkeys on Koshima islet. *Primates*, *6*(1), 1–30. https://doi.org/10.1007/BF01794457

Kim, E., Kyeong, S., Cheon, K. A., Park, B., Oh, M. K., Chun, J. W., ... & Song, D. H. (2016). Neural responses to affective and cognitive theory of mind in children and adolescents with autism spectrum disorder. *Neuroscience letters*, *621*, 117-125.

Kline, M. A. (2015). How to learn about teaching: An evolutionary framework for the study of teaching behavior in humans and other animals. *The Behavioral and Brain Sciences*, *38*(May), e31. https://doi.org/10.1017/S0140525X14000090

Kovacs, A., Teglas, E., & Endress, A. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830–1834.

Krachun, C., Call, J., & Tomasello, M. (2010). A New Change-of-Contents False Belief Test: Children and Chimpanzees Compared. *International Journal of Comparative Psychology*, *23*(2), 145–165.

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science*, *12*(4), 521–535. https://doi.org/10.1111/j.1467-7687.2008.00793.x

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

Kulke, L., Johannsen, J., & Rakoczy, H. (2019). Why can some implicit Theory of Mind tasks be replicated and others cannot? A test of mentalizing versus submentalizing accounts. *PLoS ONE*, *14*(3). https://doi.org/10.1371/journal.pone.0213772

Kulke, L., & Rakoczy, H. (2018). Implicit Theory of Mind – An overview of current replications and non-replications. *Data in Brief*, *16*, 101–104. https://doi.org/10.1016/j.dib.2017.11.016

Kulke, L., von Duhn, B., Schneider, D., & Rakoczy, H. (2018). Is Implicit Theory of Mind a Real and Robust Phenomenon? Results From a Systematic Replication Study. *Psychological Science*, *29*(6), 888–900. https://doi.org/10.1177/0956797617747090

Kummer, H., & Goodall, J. (2012). Conditions of Innovative Behaviour in Primates. *Animal Innovation*, *214*, 203–214. https://doi.org/10.1093/acprof:oso/9780198526223.003.0010

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

Laland, K. N. (2017). The origins of language in teaching. *Psychonomic Bulletin and Review*, *24*(1), 225–231. https://doi.org/10.3758/s13423-016-1077-7

Lehner, S. R., Burkart, J. M., & van Schaik, C. P. (2011). Can captive orangutans (Pongo pygmaeus abelii) be coaxed into cumulative build-up of techniques? *Journal of Comparative Psychology*, *125*(4), 446–455. https://doi.org/10.1037/a0024413

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (1.3.4).

Leslie, A. M. (2005). Developmental parallels in understanding minds and bodies. *Trends in Cognitive Sciences*, *9*(10), 459–462. https://doi.org/10.1016/j.tics.2005.08.006

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

Low, J., & Watts, J. (2013). Attributing False Beliefs About Object Identity Reveals a Signature Blind Spot in Humans' Efficient Mind-Reading System. *Psychological Science*, *24*(3), 305–311. https://doi.org/10.1177/0956797612451469

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://royalsocietypublishing.org/doi/10.1098/rspb.2020.1885

Lycett, S. J., & Von Cramon-Taubadel, N. (2013). A 3D morphometric analysis of surface geometry in Levallois cores: Patterns of stability and variability across regions and their implications. *Journal of Archaeological Science*, *40*(3), 1508–1517. https://doi.org/10.1016/j.jas.2012.11.005

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 19751–19756. https://doi.org/10.1073/pnas.0704452104

Manrique, H. M., Völter, C. J., & Call, J. (2013). Repeated innovation in great apes. *Animal Behaviour*, *85*(1), 195–202. https://doi.org/10.1016/j.anbehav.2012.10.026

Maoz, H., Tsviban, L., Gvirts, H. Z., Shamay-Tsoory, S. G., Levkovitz, Y., Watemberg, N., & Bloch, Y. (2014). Stimulants improve theory of mind in children with attention deficit/hyperactivity disorder. *Journal of Psychopharmacology*, *28*(3), 212-219.

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

Marticorena, D. C. W., Ruiz, A. M., Mukerji, C., Goddu, A., & Santos, L. R. (2011). Monkeys represent others' knowledge but not their beliefs. *Developmental Science*, *14*(6), 1406–1416. https://doi.org/10.1111/j.1467-7687.2011.01085.x

Martin, A., & Santos, L. R. (2014). The origins of belief representation: Monkeys fail to automatically represent others' beliefs. *Cognition*, *130*(3), 300–308. https://doi.org/10.1016/j.cognition.2013.11.016

Maynard, A. E. (2002). Cultural teaching: The development of teaching skills in maya sibling interactions. *Cultural Development*, *73*(3), 969–982. https://doi.org/10.4324/9781315260495-14

McGuigan, N., Makinson, J., & Whiten, A. (2011). From over-imitation to super-copying: Adults imitate causally irrelevant aspects of tool use with higher fidelity than young children. *British Journal of Psychology*, *102*(1), 1–18. https://doi.org/10.1348/000712610X493115

Melis, A. P., Call, J., & Tomasello, M. (2006). Chimpanzees (Pan troglodytes) conceal visual and auditory information from others. *Journal of Comparative Psychology*, *120*(2), 154. https://doi.org/10.1111/j.1467-7687.2009.00892.x

Mesoudi, A. (2011). An experimental comparison of human social learning strategies: Payoff-biased social learning is adaptive but underused. *Evolution and Human Behavior*, *32*(5), 334–342. https://doi.org/10.1016/j.evolhumbehav.2010.12.001

Mesoudi, A., & Thornton, A. (2018). *What is cumulative cultural evolution?* https://doi.org/10.1098/rspb.2018.0712

Mesoudi, A., Whiten, A., & Dunbar, R. (2006). A bias for social information in human cultural transmission. *British Journal of Psychology*, *97*(3), 405–423. https://doi.org/10.1348/000712605X85871

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

Muthukrishna, M., & Henrich, J. (2016). Innovation in the collective brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1690). https://doi.org/10.1098/rstb.2015.0192

Newcombe, N., & Huttenlocher, J. (1992). Children's Early Ability to Solve Perspective-Taking Problems. *Developmental Psychology*, *28*(4), 635–643. https://doi.org/10.1037/0012-1649.28.4.635

Oesch, N., & Dunbar, R. I. M. (2017). The emergence of recursion in human language: Mentalising predicts recursive syntax task performance. *Journal of Neurolinguistics*, *43*, 95–106. https://doi.org/10.1016/j.jneuroling.2016.09.008

Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, *308*(5719), 255–258. https://doi.org/10.1126/science.1107621

Osiurak, F., De Oliveira, E., Navarro, J., Lesourd, M., Claidière, N., & Reynaud, E. (2016). Physical intelligence does matter to cumulative technological culture. *Journal of Experimental Psychology: General*, *145*(8), 941–948. https://doi.org/10.1037/xge0000189

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). *PsychoPy2: experiments in behavior made easy.*

Penn, D. C., & Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a "theory of mind". *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1480), 731–744. https://doi.org/10.1098/rstb.2006.2023

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*(2), 125–137. https://doi.org/10.1111/j.2044-835x.1987.tb01048.x

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that…" attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, *39*, 437–471.

Piaget, J., & Inhelder, B. (1956). *The child's conception of space.* Routledge & Kegan Paul.

Pillow, B. H., & Flavell, J. H. (1986). Young children's knowledge about visual perception: projective size and shape. *Child Development*, *57*(1), 125–135.

https://doi.org/10.1111/j.1467-8624.1986.tb00013.x

Podos, J., & Warren, P. S. (2007). The Evolution of Geographic Variation in Birdsong. *Advances in the Study of Behavior*, *37*(07), 403–458. https://doi.org/10.1016/S0065-3454(07)37009-5

Poulin-Dubois, D., Polonia, A., & Yott, J. (2013). Is False Belief Skin-Deep? The Agent's Eye Status Influences Infants' Reasoning in Belief-Inducing Situations. *Journal of Cognition and Development*, *14*(1), 87–99. https://doi.org/10.1080/15248372.2011.608198

Poulin-Dubois, D., & Yott, J. (2017). Probing the depth of infants' theory of mind: disunity in performance across paradigms. *Developmental Science*, *21*(4), 1–11. https://doi.org/10.1111/desc.12600

Povinelli, D. J., & Vonk, J. (2004). We Don't Need a Microscope to Explore the Chimpanzee's Mind. *Mind and Language*, *19*(1), 1–28. https://doi.org/10.1111/j.1468-0017.2004.00244.x

Powell, L. J., Hobbs, K., Bardis, A., Carey, S., & Saxe, R. (2018). Replications of implicit theory of mind tasks with varying representational demands. *Cognitive Development*, *46*, 40–50. https://doi.org/10.1016/j.cogdev.2017.10.004

Premack, D. (2007). Human and animal cognition: Continuity and discontinuity. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(35), 13861–13867. https://doi.org/10.1073/pnas.0706147104

Premack, D., & Woodruff, G. (1978a). Chimpanzee problem-solving: a test for comprehension. *Science*, *202*(4367), 532–535. https://doi.org/10.1126/science.705342

Premack, D., & Woodruff, G. (1978b). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *4*(1978), 515–526.

Qualtrics. (2020). *Qualtrics*. https://www.qualtrics.com

Quesque, F., & Rossetti, Y. (2020). What Do Theory-of-Mind Tasks Actually Measure? Theory and Practice. *Perspectives on Psychological Science*, *15*(2), 384–396.

https://doi.org/10.1177/1745691619896607

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117*(2), 230–236. https://doi.org/10.1016/j.cognition.2010.08.003

Ravignani, A., Delgado, T., & Kirby, S. (2017). Musical evolution in the lab exhibits rhythmic universals. *Nature Human Behaviour*, *1*(1), 1–7. https://doi.org/10.1038/s41562-016-0007

Reader, S. M., & Laland, K. N. (2003). *Animal innovation* (10th ed.). Oxford University Press.

Reindl, E., & Tennie, C. (2018). Young children fail to generate an additive ratchet effect in an open-ended construction task. *PLoS ONE*, *13*(6), 1–22. https://doi.org/10.1371/journal.pone.0197828

Ronfard, S., & Corriveau, K. H. (2016). Journal of Experimental Child Teaching and preschoolers ' ability to infer knowledge from mistakes. *Journal of Experimental Child Psychology*, *150*, 87–98. https://doi.org/10.1016/j.jecp.2016.05.006

Ronfard, S., Was, A. M., & Harris, P. L. (2016). Children teach methods they could not discover for themselves. *Journal of Experimental Child Psychology*, *142*, 107–117. https://doi.org/10.1016/j.jecp.2015.09.032

Rossetto, F., Castelli, I., Baglio, F., Massaro, D., Alberoni, M., Nemni, R., Shamay-Tsoory, S., & Marchetti, A. (2018). Cognitive and Affective Theory of Mind in Mild Cognitive Impairment and Parkinson's Disease: Preliminary Evidence from the Italian Version of the Yoni Task. *Developmental Neuropsychology*, *43*(8), 764–780. https://doi.org/10.1080/87565641.2018.1529175

Rubio-Fernández, P., & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological Science*, *24*(1), 27–33. https://doi.org/10.1177/0956797612447819

Salatas, H., & Flavell, J. H. (1976). Perspective Taking: The Development of Two

Components of Knowledge. *Child Development*, *47*(1), 103. https://doi.org/10.2307/1128288

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255. https://doi.org/10.1128/AAC.03728-14

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. M. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 929–937. https://doi.org/10.1037/a0035175

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

Savage-Rumbaugh, E. S., Rumbaugh, D. M., & Boysen, S. (1978). Symbollic Communication Between Two Chimpanzees. *Science*, *201*(4356), 641–644.

Schneider, D., Bayliss, A. P., Becker, S. I., & Dux, P. E. (2012). Eye movements reveal sustained implicit processing of others' mental states. *Journal of Experimental Psychology. General*, *141*(3), 433–438. https://doi.org/10.1037/a0025458

Schober, F., & Clark, H. H. (1989). Understanding by Addressees and Overhearers. *Cognitive Psychology*, *21*, 211–232.

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

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind Eyes : An Absence of Asperger Syndrome. *Science*, *219*(August), 883–885. https://doi.org/10.1126/science.1176170

Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: A lesion study. *Neuropsychologia*, *45*(13), 3054–3067. https://doi.org/10.1016/j.neuropsychologia.2007.05.021

Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C. M., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*, *18*(4), 186–193. https://doi.org/10.1016/j.tics.2014.01.006

Smith, J. W., Benkman, C. W., & Coffey, K. (1999). The use and misuse of public information by foraging red crossbills. *Behavioral Ecology*, *10*(1), 54–62. https://doi.org/10.1093/beheco/10.1.54

Sol, D., Duncan, R. P., Blackburn, T. M., Cassey, P., & Lefebvre, L. (2005). Big brains, enhanced cognition, and response of birds to novel environments. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(15), 5460–5465. https://doi.org/10.1073/pnas.0408145102

Sol, D., Lefebvre, L., & Rodríguez-Teijeiro, J. D. (2005). Brain size, innovative propensity and migratory behaviour in temperate Palaearctic birds. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1571), 1433–1441. https://doi.org/10.1098/rspb.2005.3099

Song, H. joo, Onishi, K. H., Baillargeon, R., & Fisher, C. (2008). Can an agent’s false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition*, *109*(3), 295–315. https://doi.org/10.1016/j.cognition.2008.08.008

Southgate, V., Senju, A., & Csibra, G. (2007). Action Anticipation Through Attribution of False Belief by 2-Year-Olds. *Psychological Science*, *18*(7), 587–592.

Spaulding, S. (2020). What is mindreading? *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(3), 1–13. https://doi.org/10.1002/wcs.1523

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*(4), 605–632.

Strauss, S., Ziv, M., & Stein, A. (2002). Teaching as a natural cognition and its relations to preschoolers’ developing theory of mind. *Cognitive Development*, *17*(3–4), 1473–1487. https://doi.org/10.1016/S0885-2014(02)00128-4

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers Can Attribute Second-Order Beliefs. *Developmental Psychology*, *30*(3), 395–402. https://doi.org/10.1037/0012-1649.30.3.395

Surian, L, Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13- month-old infants. *Psychological Science*, *18*(7), 580–586.

Surian, L. & Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *British Journal of Developmental Psychology*, *30*(1), 30–44. https://doi.org/10.1111/j.2044-835X.2011.02046.x

Tamariz, M., & Kirby, S. (2015). Culture: Copying, compression, and conventionality. *Cognitive Science*, *39*(1), 171–183. https://doi.org/10.1111/cogs.12144

Taylor, M., Esbensen, B. M., & Bennett, R. T. (1994). Children's Understanding of Knowledge Acquisition: The Tendency for Children to Report That They Have Always Known What They Have Just Learned. *Child Development*, *65*(6), 1581. https://doi.org/10.2307/1131282

Team, R. C. (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.r-project.org/

Team, Rs. (2020). *RStudio: Integrated Development for R.* http://www.rstudio.com/

Templeton, J. J., & Giraldeau, L. A. (1996). Vicarious sampling: The use of personal and public information by starlings foraging in a simple patchy environment. *Behavioral Ecology and Sociobiology*, *38*(2), 105–114. https://doi.org/10.1007/s002650050223

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

Thornton, A., & MaCuliffe, K. (2006). Teaching in Wild Meerkats. *Science*, *313*(5784), 227–229. https://doi.org/10.1126/science.1128727

Thornton, A., & Raihani, N. J. (2008). The evolution of teaching. *Animal Behaviour*, *75*(6), 1823–1836. https://doi.org/10.1111/j.1558-5646.2011.01370.x

Tomasello, M. (1999). 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., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(05), 675–735. https://doi.org/10.1017/S0140525X05000129

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences*, *16*(3), 495–552. https://doi.org/10.1017/s0140525x00031496

Topál, J., Byrne, R. W., Miklósi, Á., & Csányi, V. (2006). Reproducing human actions and action sequences: "Do as I do!" in a dog. *Animal Cognition*, *9*(4), 355–367. https://doi.org/10.1007/s10071-006-0051-6

Träuble, B., Marinović, V., & Pauen, S. (2010). Early theory of mind competencies: Do infants understand others' beliefs? *Infancy*, *15*(4), 434–444. https://doi.org/10.1111/j.1532-7078.2009.00025.x

Uller, C., & Nichols, S. (2000). Goal attribution in chimpanzees. *Cognition*, *76*(2), 27–34. https://doi.org/10.1016/S0010-0277(00)00078-0

Vaesen, K. (2012). The cognitive bases of human tool use. *Behavioral and Brain Sciences*, *35*(04), 203–218. https://doi.org/10.1017/S0140525X11001452

Vale, G. L., Davis, S. J., Lambeth, S. P., Schapiro, S. J., & Whiten, A. (2017). Acquisition of a socially learned tool use sequence in chimpanzees: Implications for cumulative culture. *Evolution and Human Behavior*, *38*(5), 635–644. https://doi.org/10.1016/j.evolhumbehav.2017.04.007

Valle, A., Massaro, D., Castelli, I., & Marchetti, A. (2015). Theory of Mind Development in Adolescence and Early Adulthood: The Growing Complexity of Recursive Thinking Ability. *Europe's Journal of Psychology*, *11*(1), 112–124. https://doi.org/10.5964/ejop.v11i1.829

van der Vaart, E., & Hemelrijk, C. K. (2014). 'Theory of mind' in animals: ways to make progress. *Synthese*, *191*(3), 335–354. https://doi.org/10.1007/s11229-012-0170-3

Weber, E. U., Blais, A.-R., & Betz, N. E. (2002). A domain-specific risk-attitude scale:

measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, *15*(4), 263–290. https://doi.org/10.1002/bdm.414

Wheeler, B. C. (2009). Monkeys crying wolf? Tufted capuchin monkeys use anti-predator calls to usurp resources from conspecifics. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1669), 3013–3018. https://doi.org/10.1098/rspb.2009.0544

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., Horner, V., & Marshall-Pescini, S. (2003). Cultural Panthropology. *Evolutionary Anthropology*, *12*(2), 92–105. https://doi.org/10.1002/evan.10107

Whiten, A., & Van Schaik, C. P. (2007). The evolution of animal “cultures” and social intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 603–620. https://doi.org/10.1098/rstb.2006.1998

Wilks, C. E. H., Rafetseder, E., Renner, E., Atkinson, M., & Caldwell, C. A. (In Press). Cognitive Prerequisites for Cumulative Culture are Context-Dependent: Children’s Potential for Ratcheting Depends on Cue Longevity. *J Exp Child Psychol*.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children’s understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Wood, D., Wood, H., Ainsworth, S., & O’Malley, C. (1995). *On Becoming a Tutor: Toward an Ontogenetic Model*. *13*(4), 565–581. https://doi.org/10.1207/s1532690xci1304

Yott, J., & Poulin-Dubois, D. (2016). Are Infants’ Theory-of-Mind Abilities Well Integrated? Implicit Understanding of Intentions, Desires, and Beliefs. *Journal of Cognition and Development*, *17*(5), 683–698. https://doi.org/10.1080/15248372.2015.1086771

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, *1*(1), 297–301.

https://doi.org/10.1038/nprot.2006.46

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*(1), 37–63. https://doi.org/10.1016/S0885-2014(96)90027-1

Ziv, M., & Frye, D. (2004). Children’s understanding of teaching: The role of knowledge and belief. *Cognitive Development*, *19*(4), 457–477. https://doi.org/10.1016/j.cogdev.2004.09.002

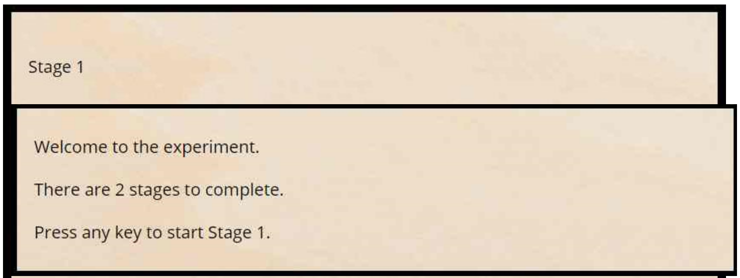
Ziv, M., Solomon, A., & Frye, D. (2008). Young children’s recognition of the intentionality of teaching. *Child Development*, *79*(5), 1237–1256. https://doi.org/10.1111/j.1467-8624.2008.01186.x

Ziv, M., Solomon, A., Strauss, S., & Frye, D. (2016). Relations Between the Development of Teaching and Theory of Mind in Early Childhood. *Journal of Cognition and Development*, *17*(2), 264–284. https://doi.org/10.1080/15248372.2015.1048862

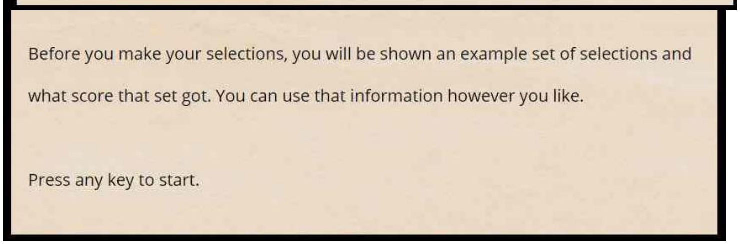
Zwirner, E., & Thornton, A. (2015). Cognitive requirements of cumulative culture: teaching is useful but not essential. *Scientific Reports*, *5*, 16781. https://doi.org/10.1038/srep16781

Appendices

Appendix 1: Chapter 2 task layout



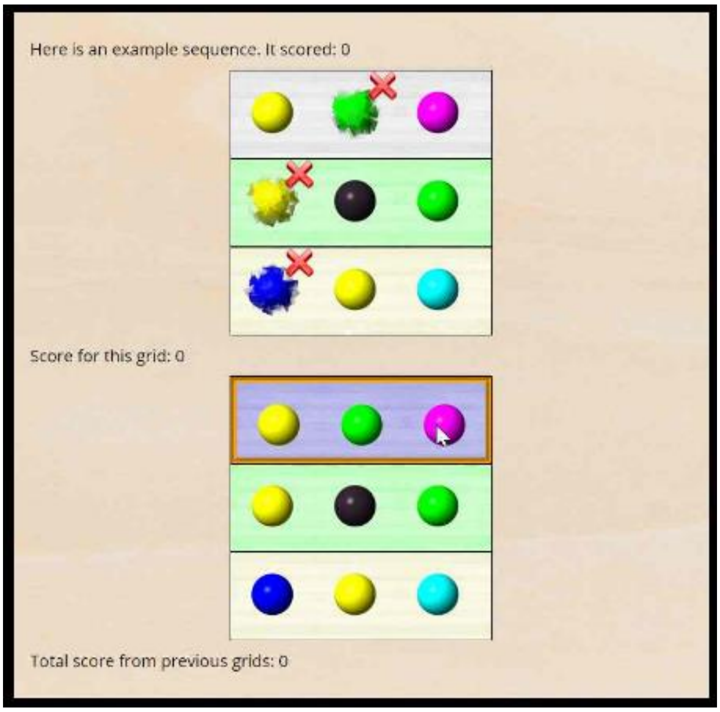
Opening screen:



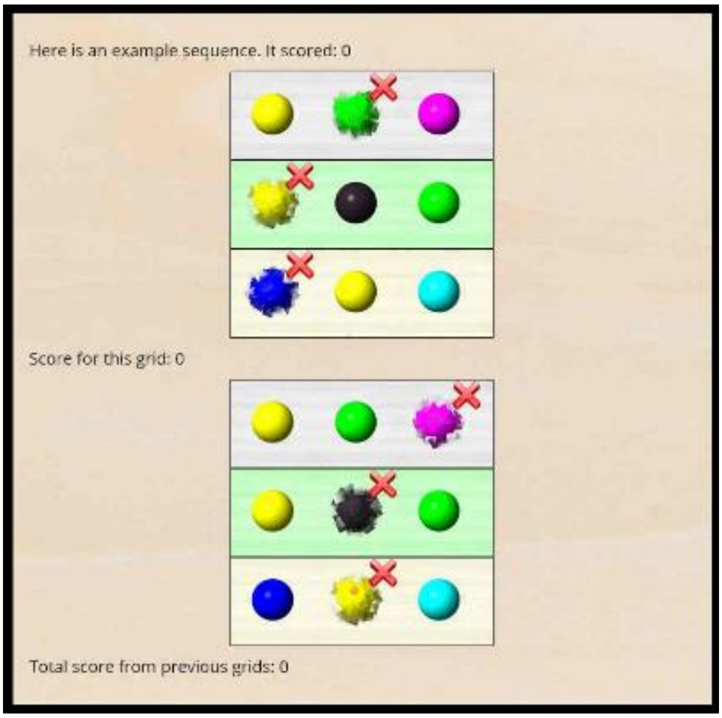
Instruction screen for stage 1:

Following the instruction screen, participants are given 88 grids showing high-precision payoff data, each with a blank grid underneath as pictured in the example below.

Demo 1: Before current participant selections



Demo 2: After current participant selections



After all 88 grids have been completed, participants get an ‘end of block screen and some instructions for block 2:

End of Block 1!

Press any key to continue.

Instruction screen for stage 2

Stage 2

You will see a series of grids, each of which contains rows of coloured balls.

One ball from each row will be worth 1 point; all others 0.

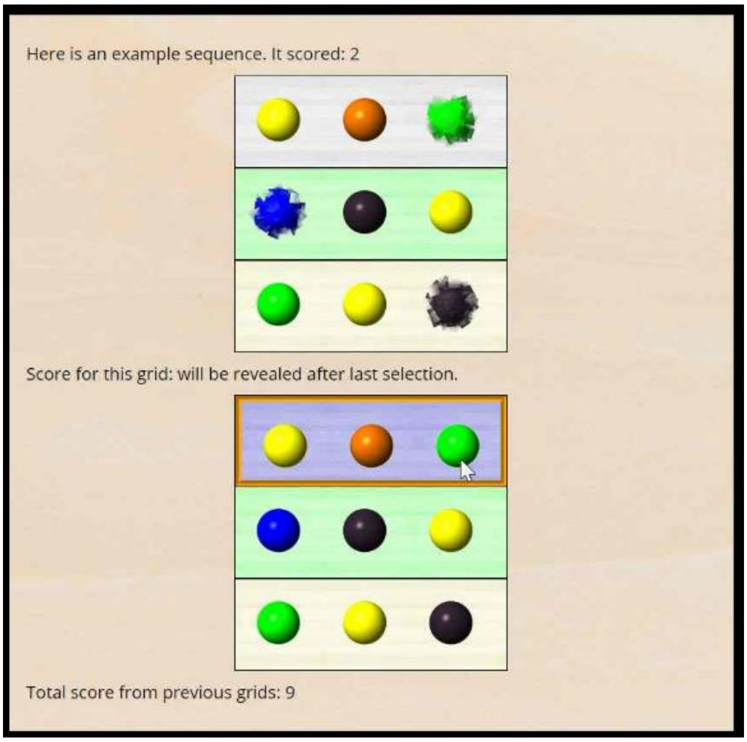
You goal is to select one ball from each row, and get as high a total score as possible.

Before you make your selections, you will be shown an example set of selections and what score that set got. You can use that information however you like.

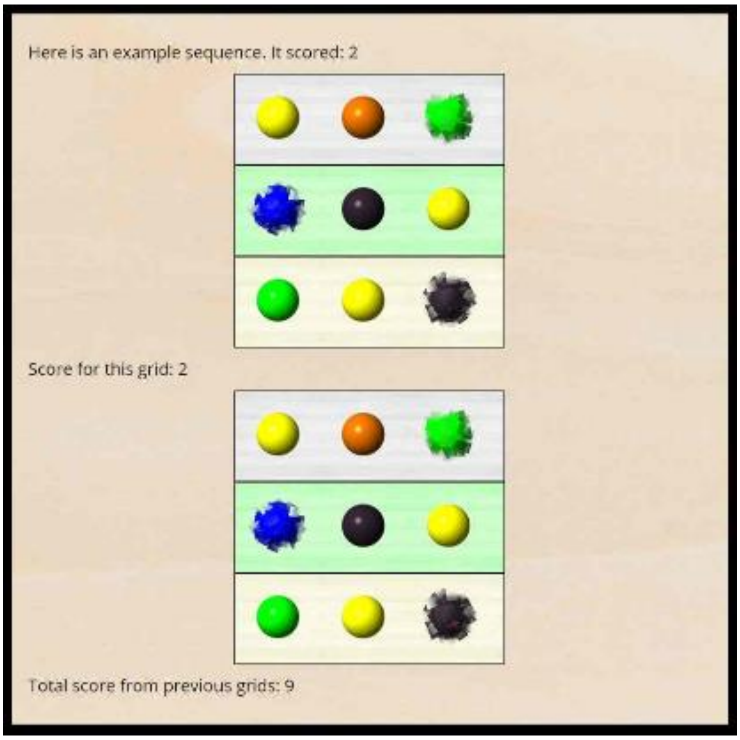
Press any key to start.

Following the instruction screen, participants are given 88 grids showing high-precision payoff data, each with a blank grid underneath as pictured in the example below.

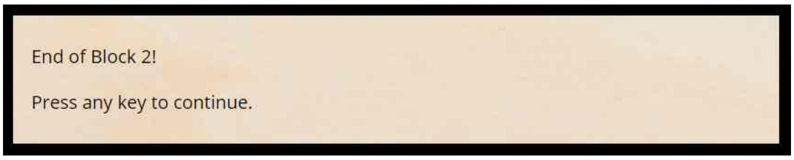
Demo 3: Before current participant selections



Demo 4: After current participant selections



After all 88 grids have been completed, participants get an ‘end of block screen and the task ends:



Appendix 2: Risk behaviour scale used in Chapter 2

This questionnaire was adapter from (Weber et al., 2002)

For each of the following statements, please indicate your likelihood of engaging in each activity or behaviour.

Provide a rating from 1 to 5 using the following scale:

0	1	2	3	4	5
Prefer not to answer	Very unlikely	Unlikely	Not sure	Likely	Very likely

1. Admitting that your tastes are different to your friend's. _____
2. Going camping in the wilderness, beyond the civilization of a campground. _____
3. Betting a day's income at the horse races. _____
4. Buying an illegal drug for your own use. _____
5. Cheating on an exam. _____
6. Chasing a tornado or hurricane by car to take dramatic photos. _____
7. Investing 10% of your annual income in a moderate growth mutual fund. _____
8. Consuming five or more servings of alcohol in a single evening. _____
9. Cheating by a significant amount on your income tax return. _____
10. Disagreeing with your father on a major issue. _____
11. Betting a day's income at a high-stake poker game. _____
12. Having an affair with a married man or woman. _____
13. Forging someone's signature. _____
14. Passing off somebody else's work as your own. _____
15. Going on vacation in a third-world country without prearranged travel and hotel accommodations. _____

16. Arguing with a friend about an issue on which he or she has a very different opinion. _____
17. Going down a ski run that is beyond your ability or closed. _____
18. Investing 5% of your annual income in a very speculative stock. _____
19. Approaching your boss to as for a raise. _____
20. Illegally copying a piece of software. _____
21. Going white water rafting during rapid water flows in the spring. _____
22. Betting a day's income on the outcome of a sporting event. _____
23. Telling a friend if his or her significant other has made a pass at you. _____
24. Investing 5% of your annual income in a conservative stock. _____
25. Shoplifting a small item (e.g., lipstick or pen). _____
26. Wearing provocative or unconventional clothing on occasion. _____
27. Engaging in unprotective sex. _____
28. Stealing an additional TV cable connection off the one you pay for. _____
29. Not wearing a seatbelt when being a passenger in the front seat. _____
30. Investing 10% of your annual income in government bonds. _____
31. Periodically engaging in a dangerous sport (e.g., mountain climbing or sky diving). _____
32. Not wearing a helmet when riding a motorcycle. _____
33. Gambling a week's income at a casino. _____
34. Taking a job that you enjoy over one that is prestigious but less enjoyable. _____
35. Defending an unpopular issue that you believe in at a social occasion. _____
36. Exposing yourself to the sun without sunscreen. _____
37. Trying out bungee jumping at least once. _____
38. Piloting your own small plane, if you could. _____
39. Walking home alone at night in a somewhat unsafe area of town. _____
40. Regularly eating high-cholesterol foods. _____

Appendix 3: Encouraging prompts for in-task, Chapter 4

GENERAL

"Remember, you get 1 point for every hit you find, and you lose a life for every miss you find!"

"These are some things the last explorer found"

INTENTIONAL ONLY

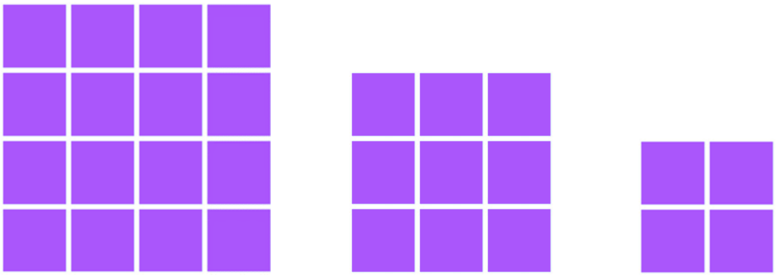
"Now you can show 3 of the places you looked to the next explorer – help the next explorer to find Alien Space Ships!"

PROMPTS

"Find the Alien Space Ships"

Appendix 4: Memory quiz used in Chapter 4

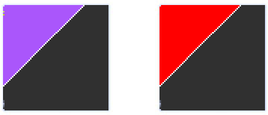
Which one is the Alien Space Ship?



How many Alien Space Ships are hiding in each universe?

1 **2** **3** **4**

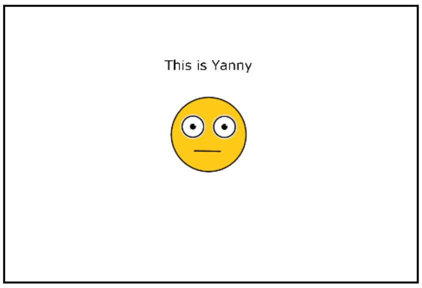
The last explorer has sent you some clues! What do they mean? Match!

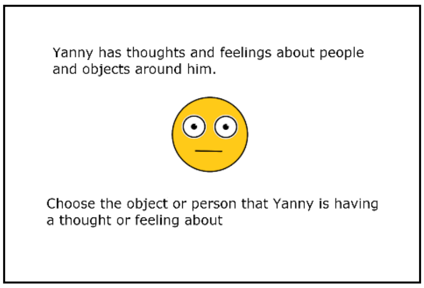


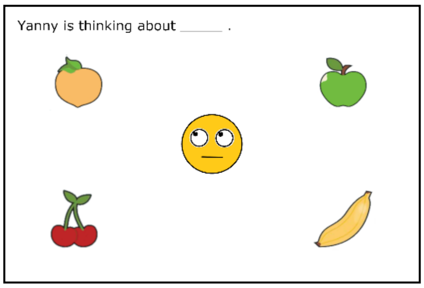
"There is a **hit** here!"

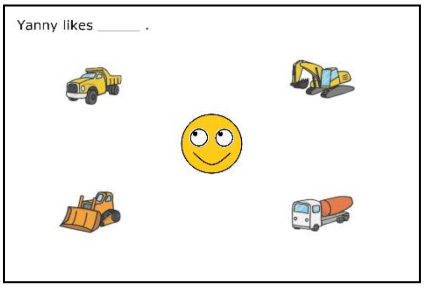
"There is **nothing** here!"

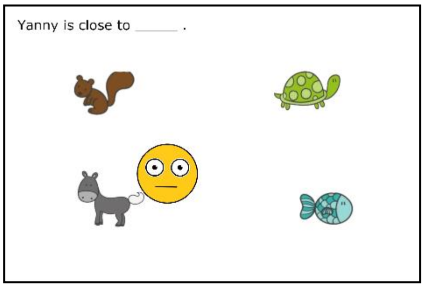
Appendix 5: Stimuli used in the theory of mind test in Chapter 4 (first-order)

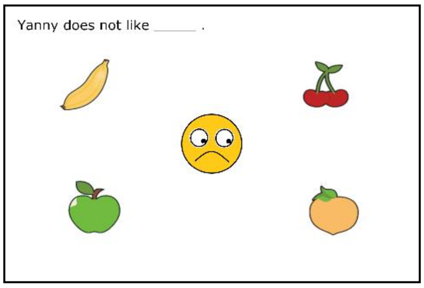


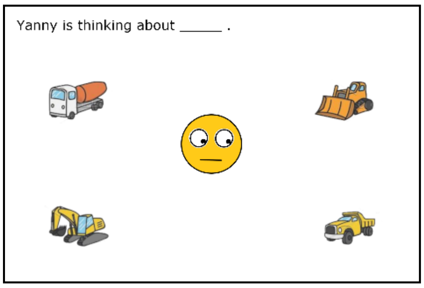


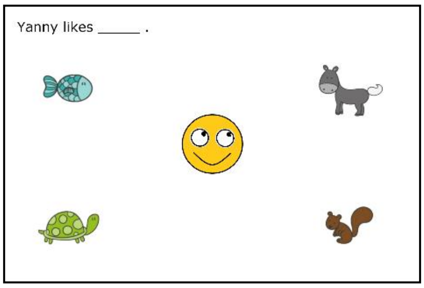


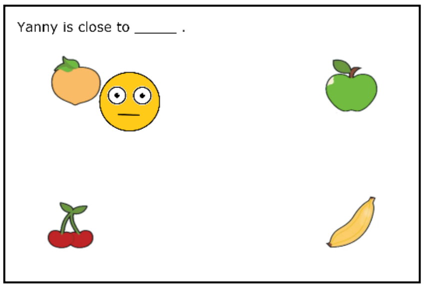


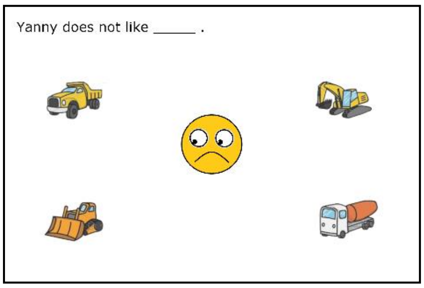


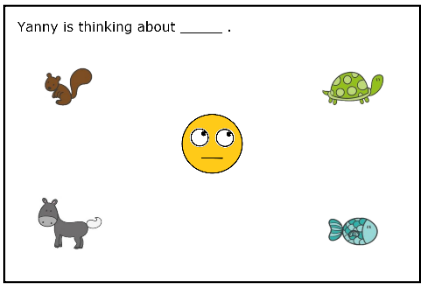


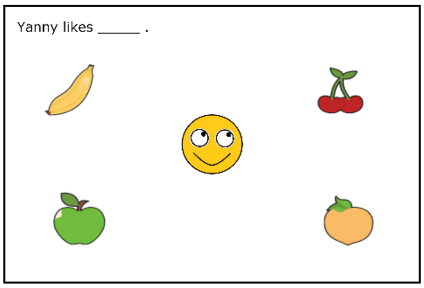


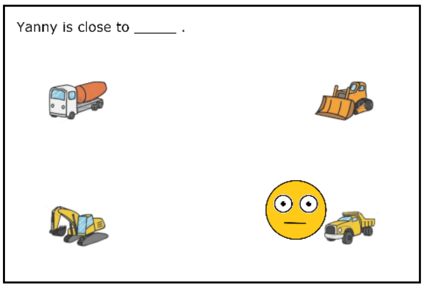


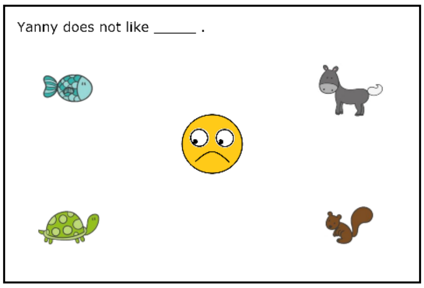




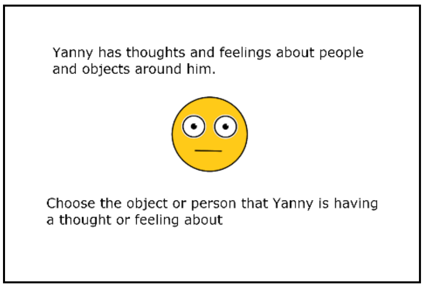


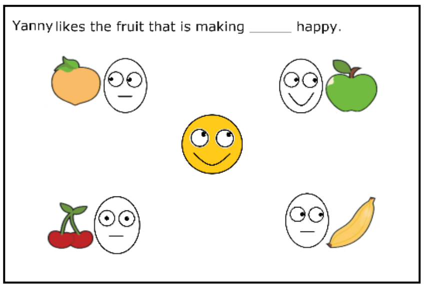


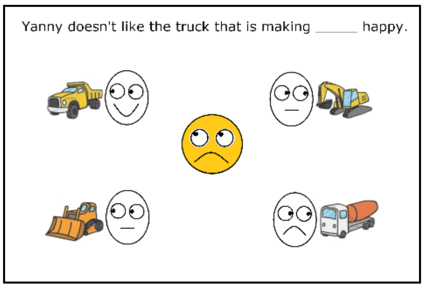

Yanny is close to ______ .

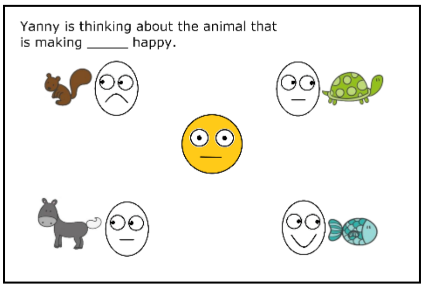

Yanny does not like ______ .

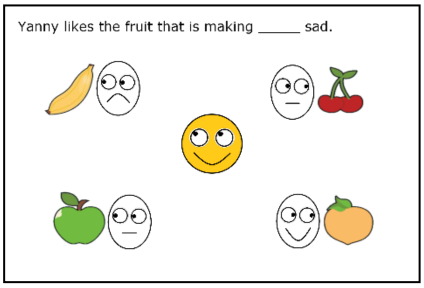
Appendix 6: Stimuli used in the theory of mind test in Chapter 4 (second-order)

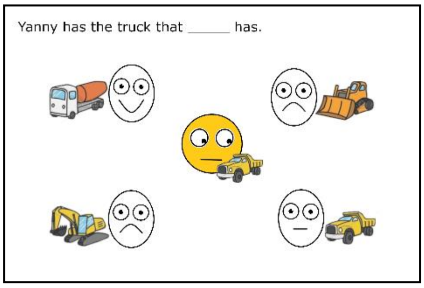


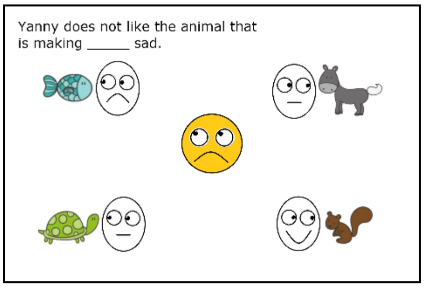


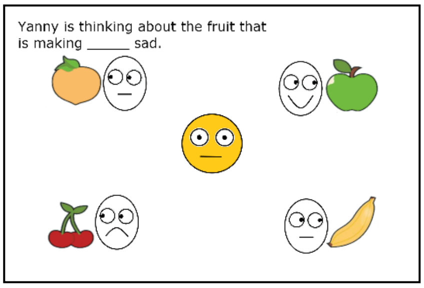


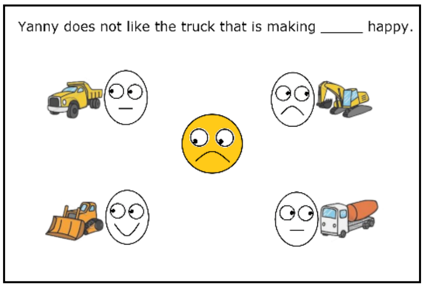


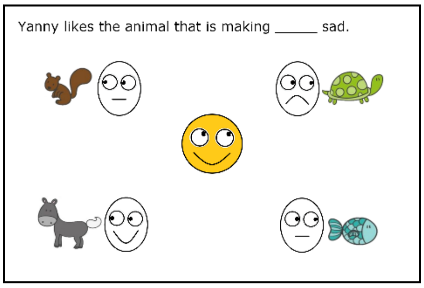


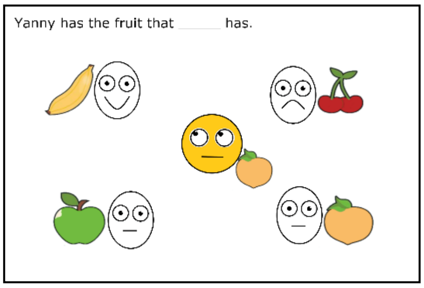


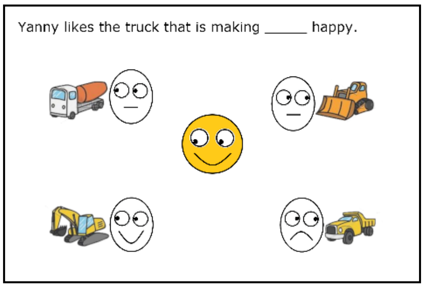


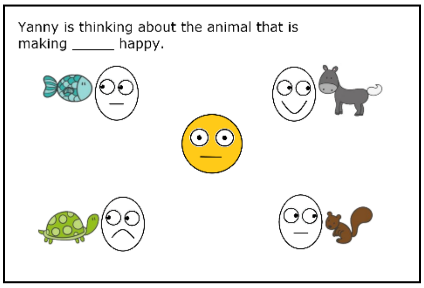




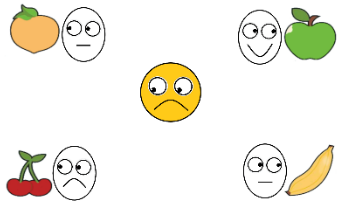




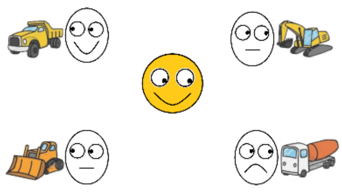




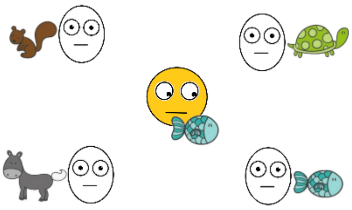
Yanny does not like the fruit that is making ______ sad.



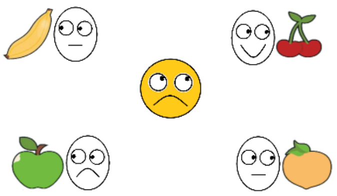
Yanny likes the truck that is making ______ sad.



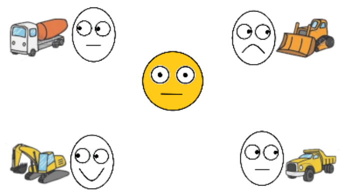
Yanny has the animal that ______ has.



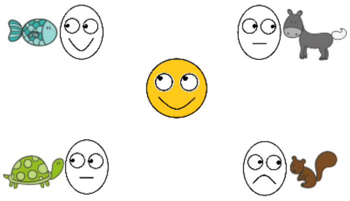
Yanny does not like the fruit that is making ______ happy.



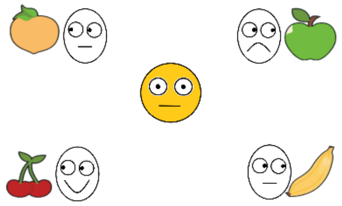
Yanny is thinking about the truck that is making ______ sad.



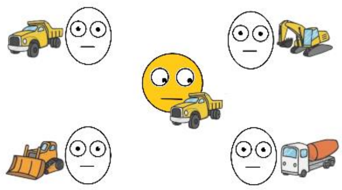
Yanny likes the animal that is making ______ happy .



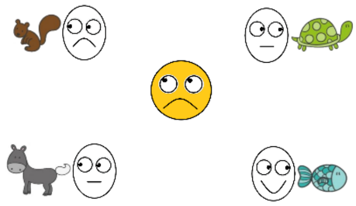
Yanny is thinking about the fruit that is making ______ happy.



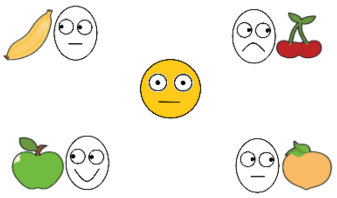
Yanny has the truck that ______ has.



Yanny does not like the animal that is making ______ sad.



Yanny is thinking about the fruit that is making ______ sad.

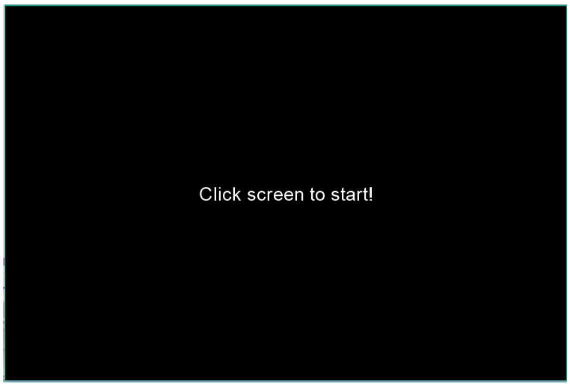


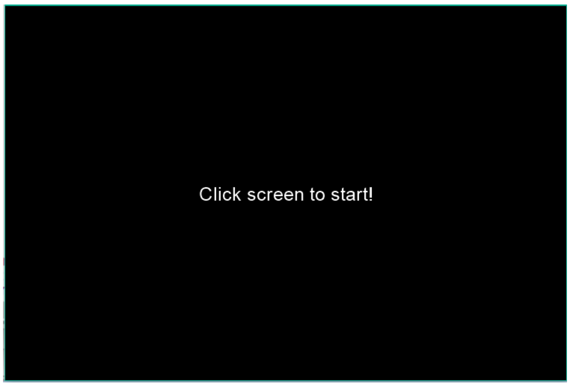
Appendix 7: Full task layout of the battleships task used with the adult sample and the child sample.

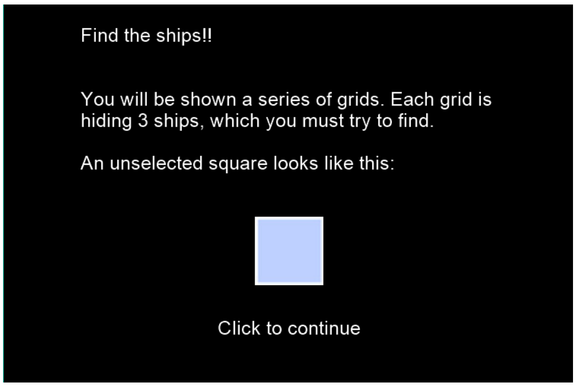
This example shows participants completing one grid, however, in the real task, adults completed ten and children completed five.

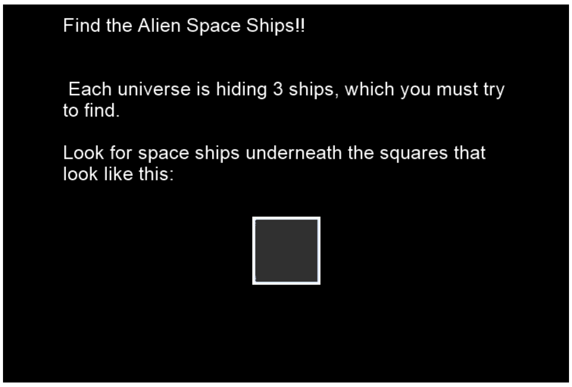
<u>Adult's Task (Chapter 3)</u> <u>Children's Task (Chapter 4)</u>

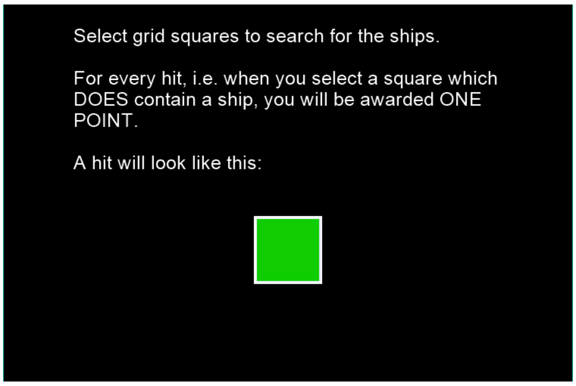
Chain position 1:

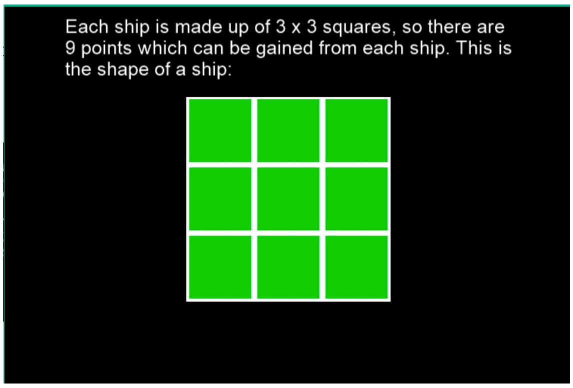


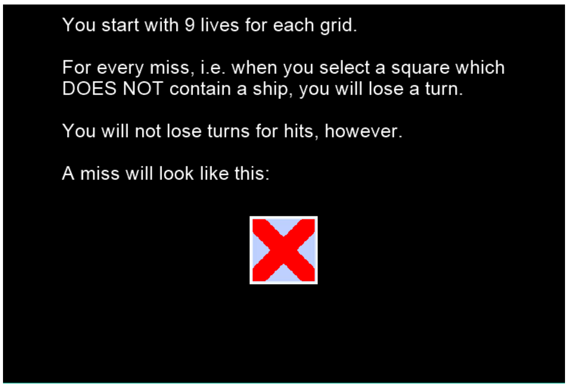




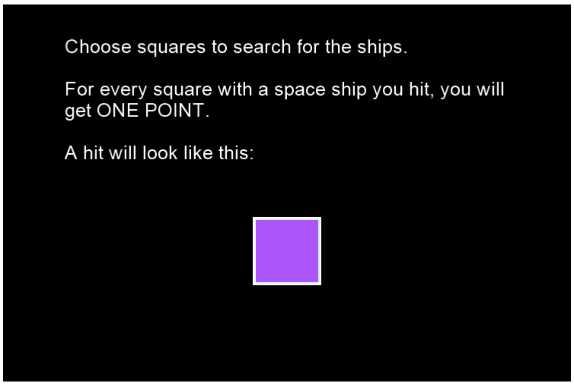


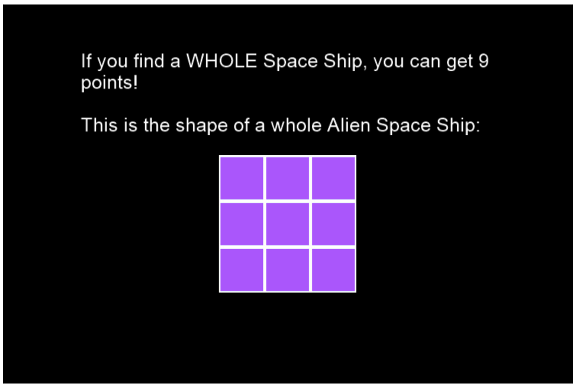

Select grid squares to search for the ships.
For every hit, i.e. when you select a square which DOES contain a ship, you will be awarded ONE POINT.
A hit will look like this:

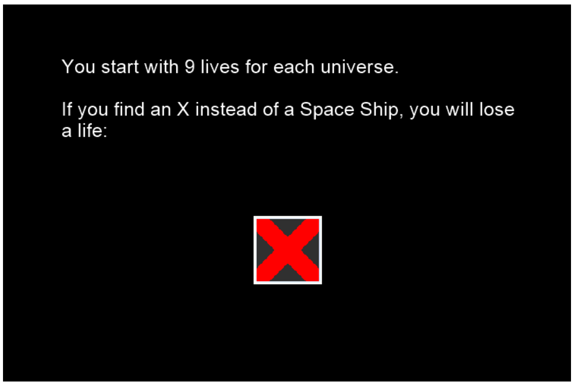

Each ship is made up of 3 x 3 squares, so there are 9 points which can be gained from each ship. This is the shape of a ship:

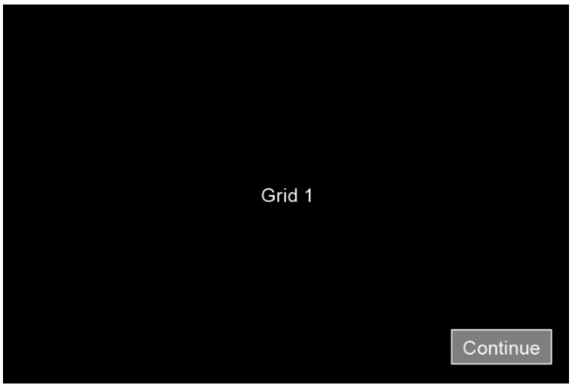

You start with 9 lives for each grid.
For every miss, i.e. when you select a square which DOES NOT contain a ship, you will lose a turn.
You will not lose turns for hits, however.
A miss will look like this:

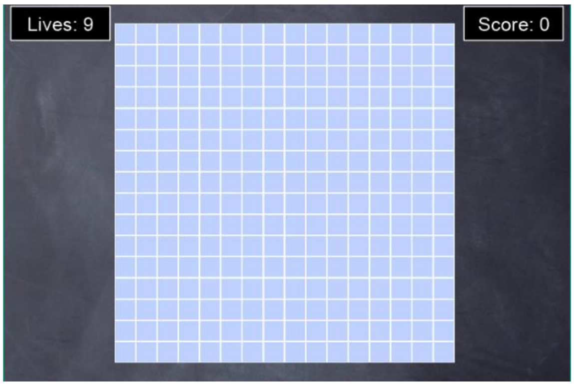

Grid 1
Continue

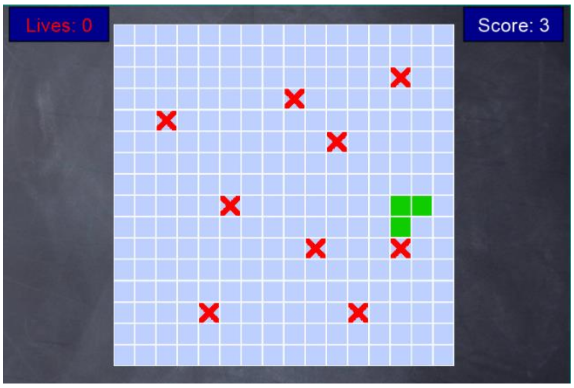

Choose squares to search for the ships.
For every square with a space ship you hit, you will get ONE POINT.
A hit will look like this:

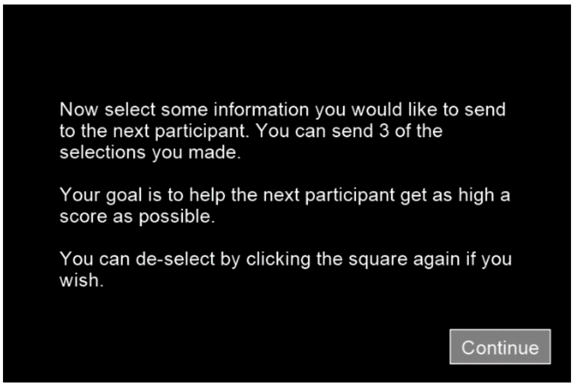

If you find a WHOLE Space Ship, you can get 9 points!
This is the shape of a whole Alien Space Ship:

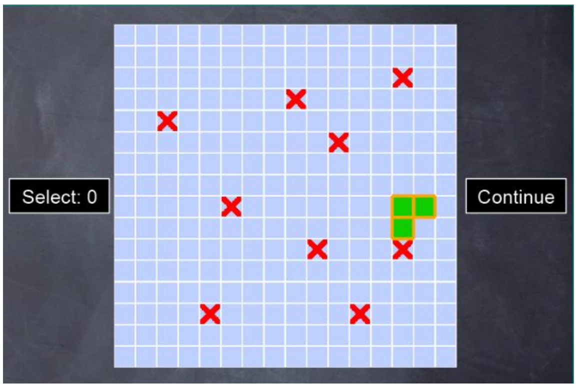

You start with 9 lives for each universe.
If you find an X instead of a Space Ship, you will lose a life:

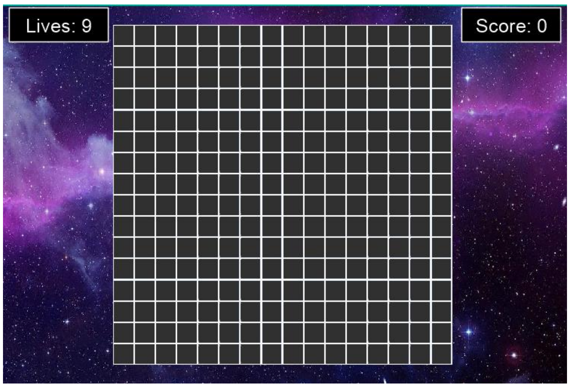

Grid 1
Continue

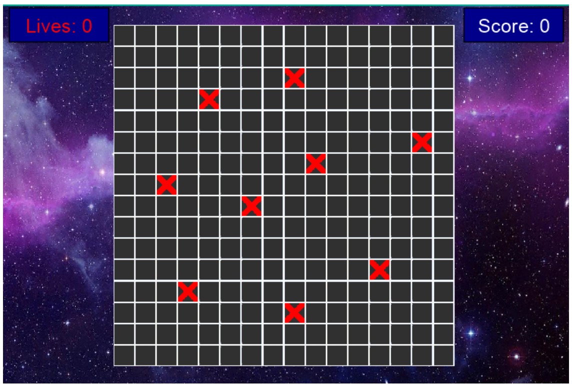

Lives: 9
Score: 0

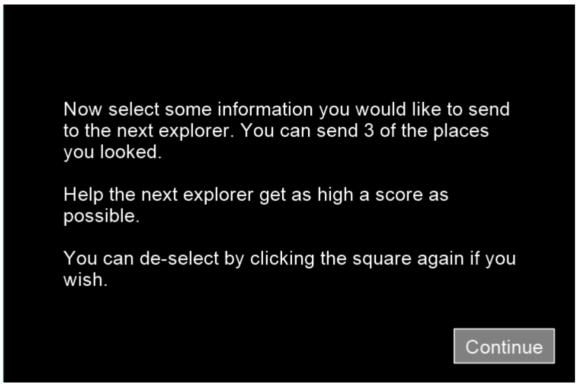

Lives: 0
Score: 3

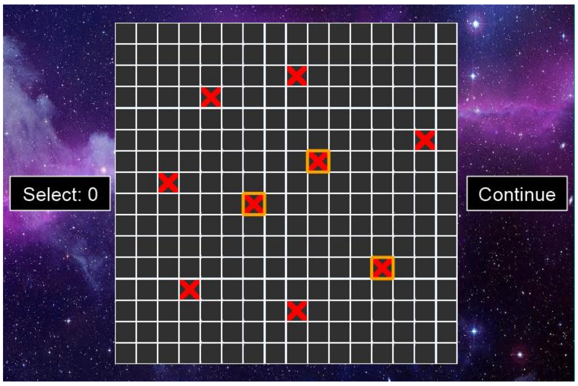

Now select some information you would like to send to the next participant. You can send 3 of the selections you made.
Your goal is to help the next participant get as high a score as possible.
You can de-select by clicking the square again if you wish.
Continue

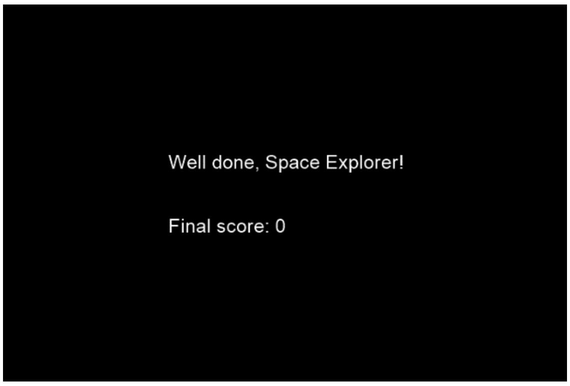

Select: 0
Continue


Lives: 9
Score: 0


Lives: 0
Score: 0

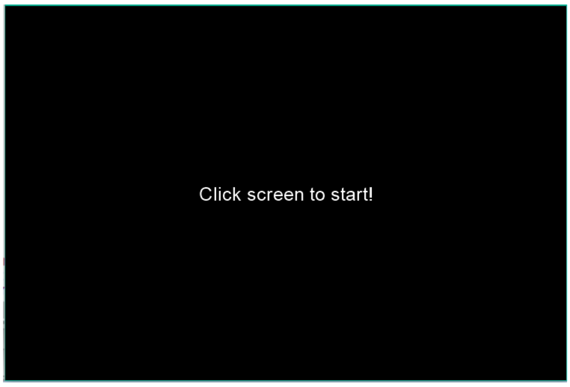

Now select some information you would like to send to the next explorer. You can send 3 of the places you looked.
Help the next explorer get as high a score as possible.
You can de-select by clicking the square again if you wish.
Continue

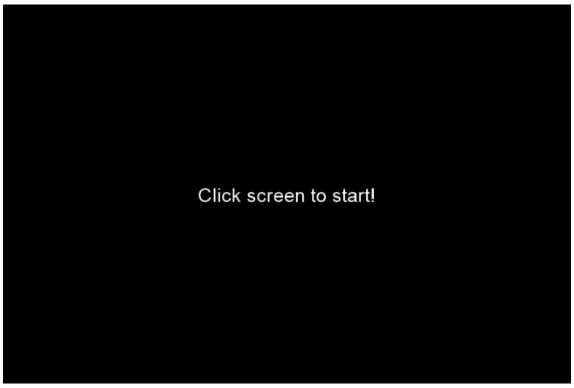

Select: 0
Continue



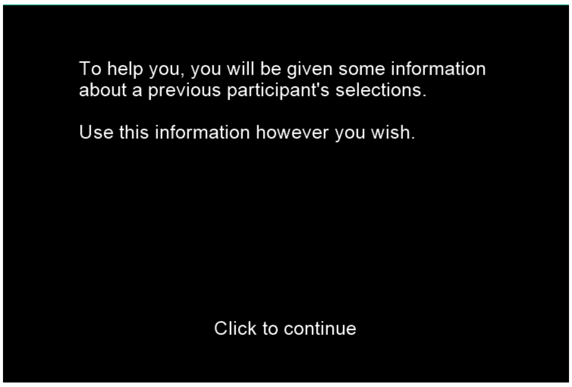
Chain position 2:

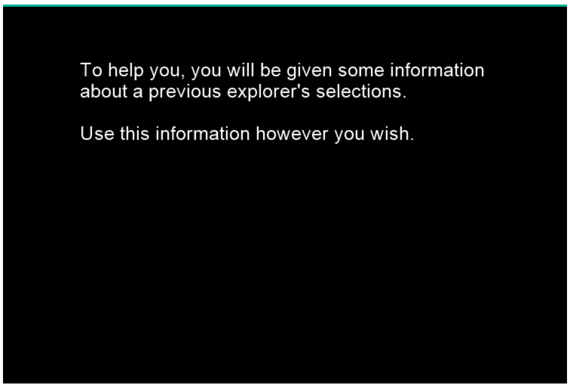
Instructions were the same as chain position 1, but with additional elements:

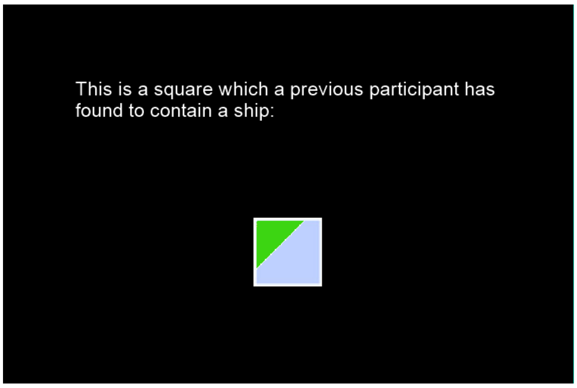


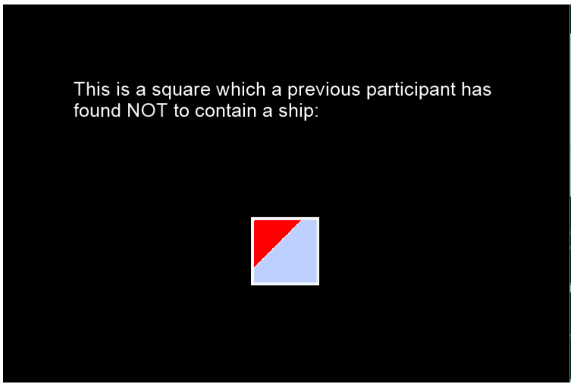


Additional instructions given to all participants after the first chain position:

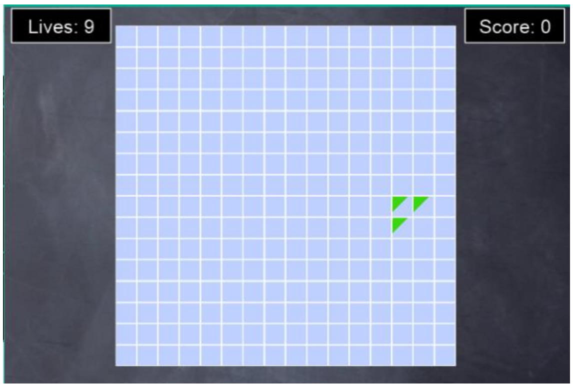


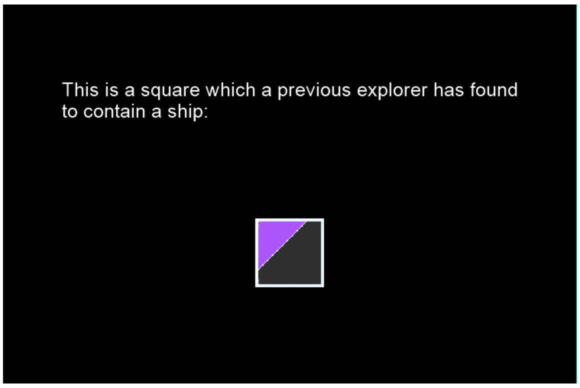


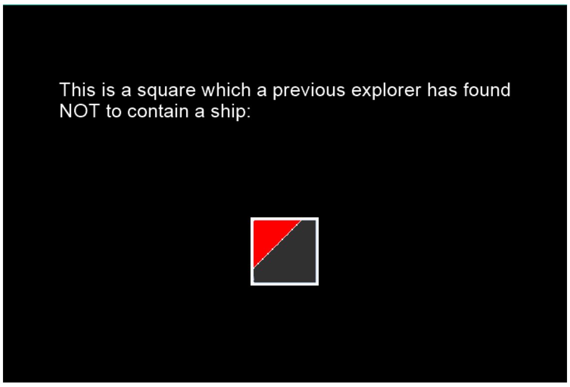

This is a square which a previous participant has found to contain a ship:


This is a square which a previous participant has found NOT to contain a ship:

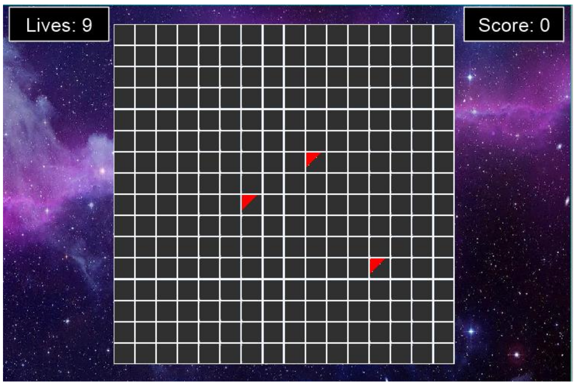

Grid 1
Continue

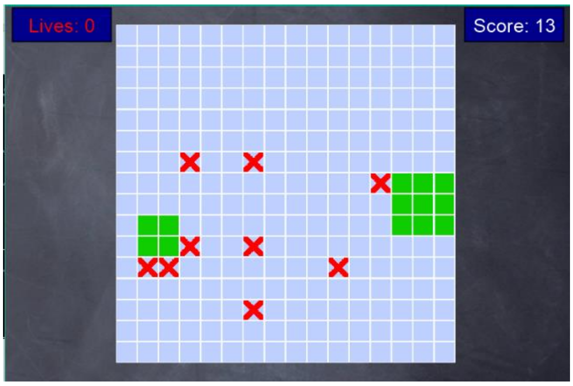

Lives: 9
Score: 0

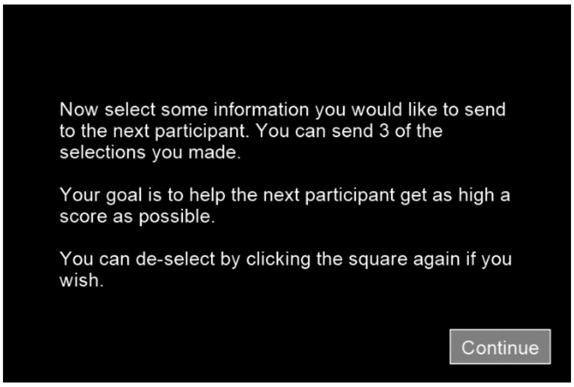

This is a square which a previous explorer has found to contain a ship:

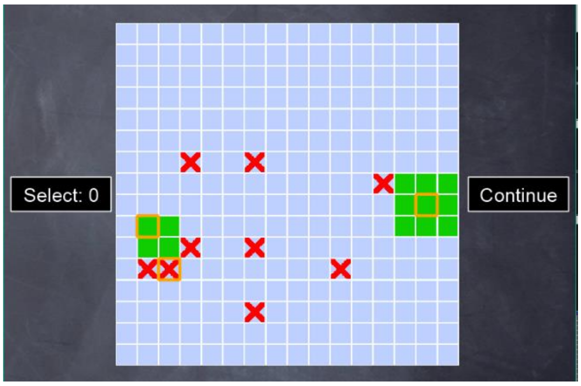

This is a square which a previous explorer has found NOT to contain a ship:

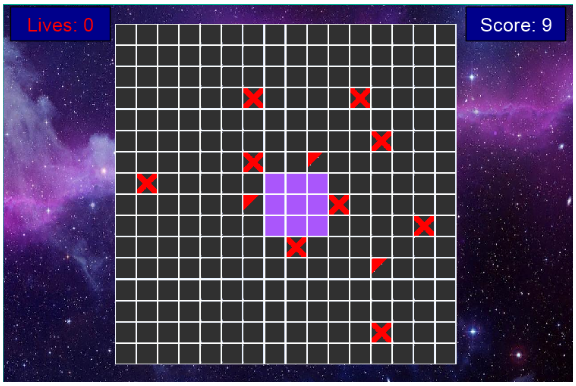

Grid 1
Continue

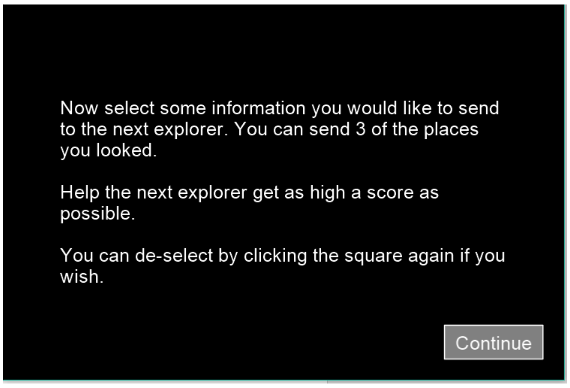

Lives: 9
Score: 0

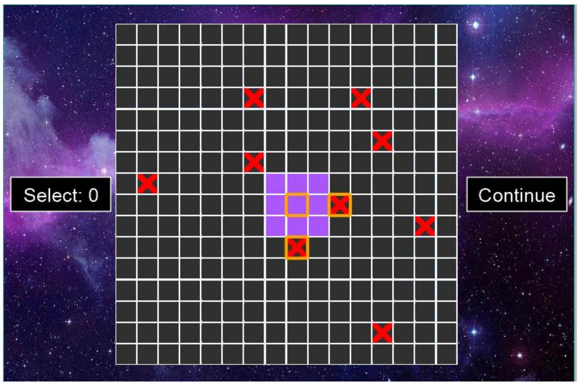


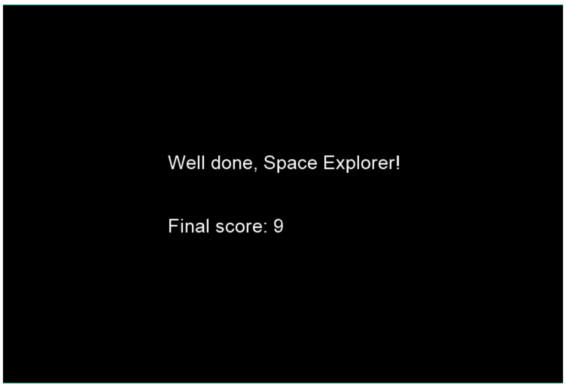






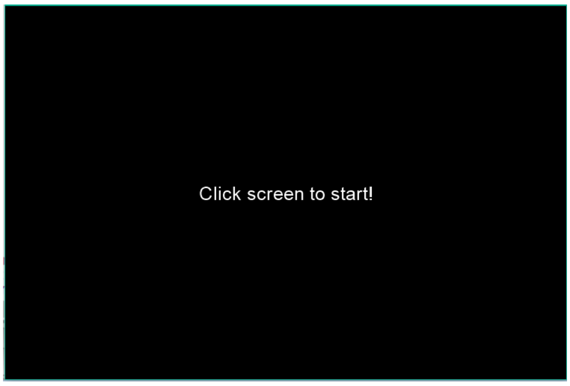


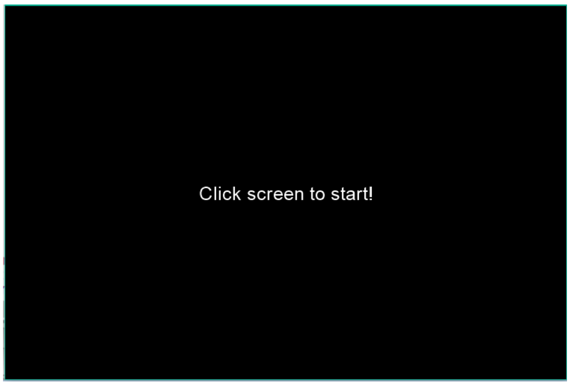


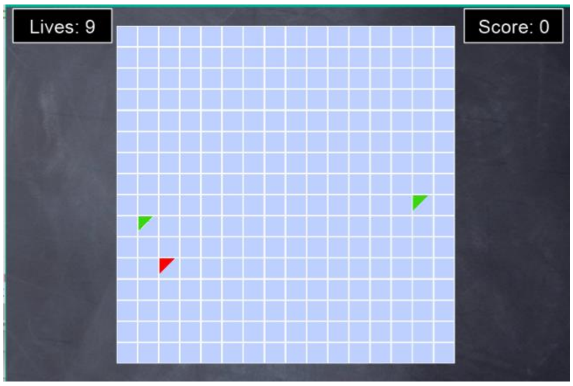


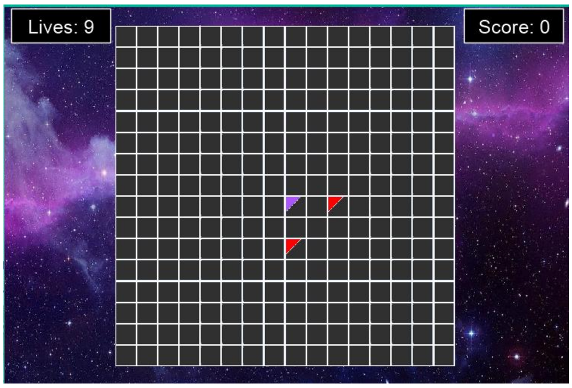
Chain position 3:

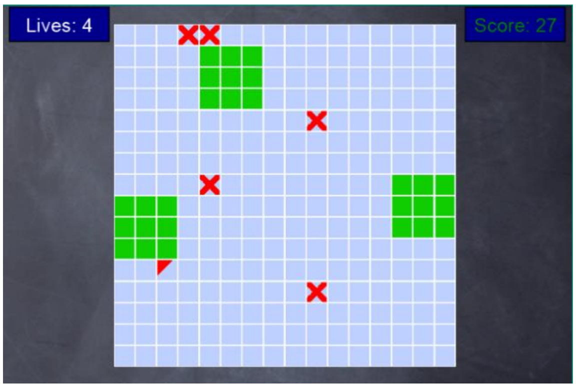
Instructions were the same as chain position 2

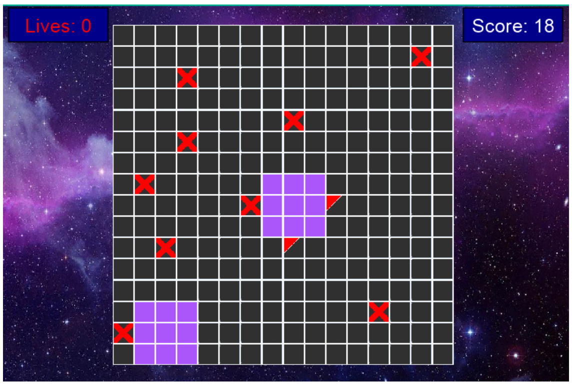


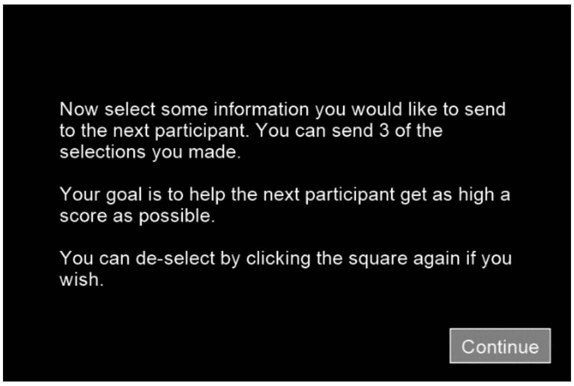


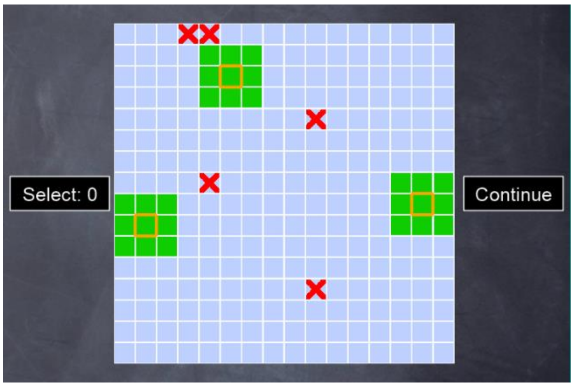


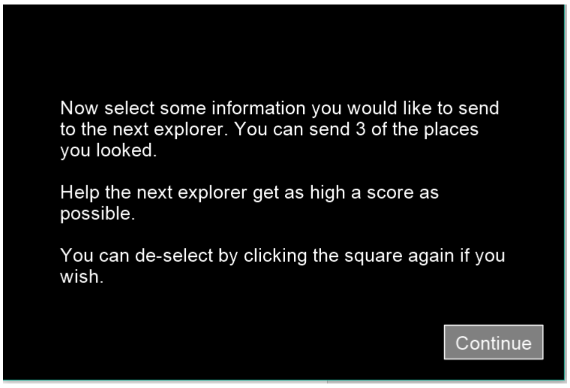


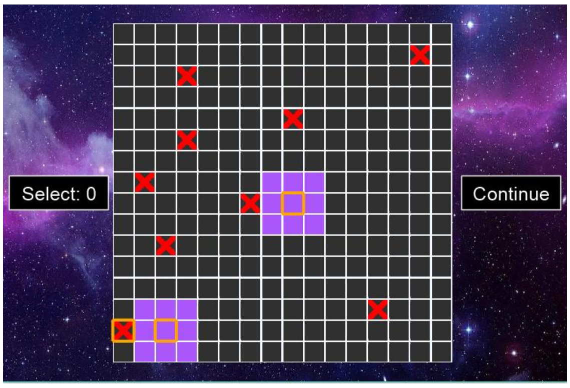








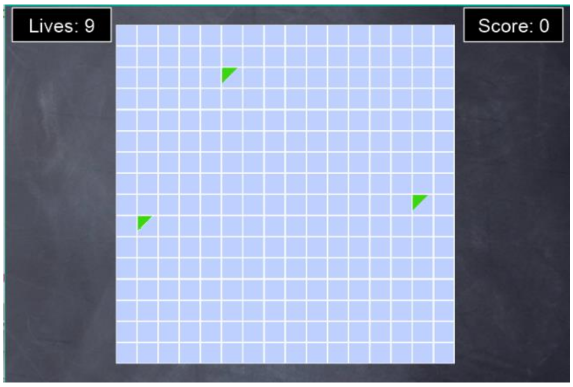


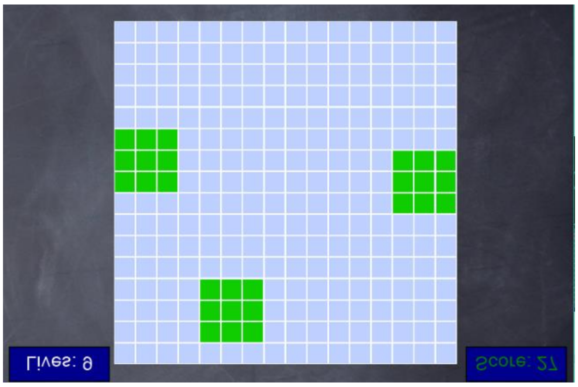


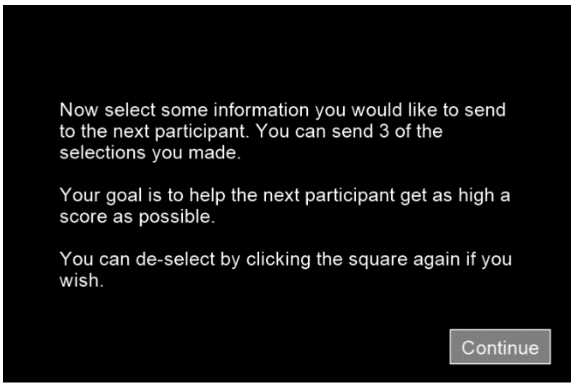
Chain position 4:

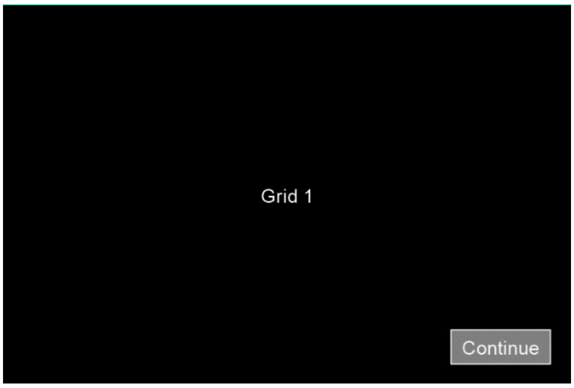
Instructions were the same as chain position 2

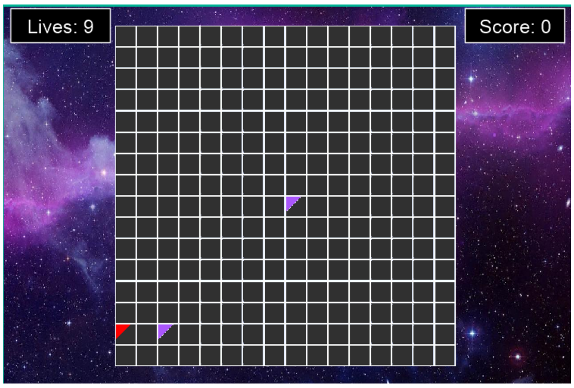

Grid 1
Continue

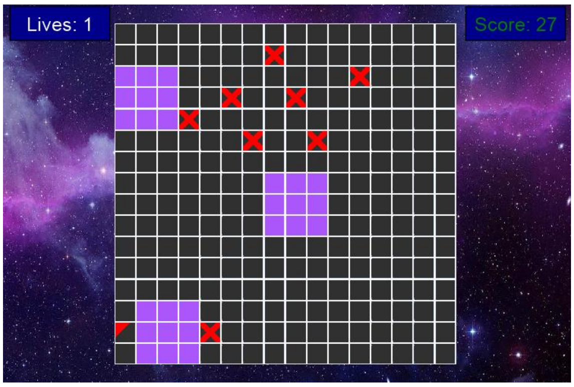

Lives: 9
Score: 0

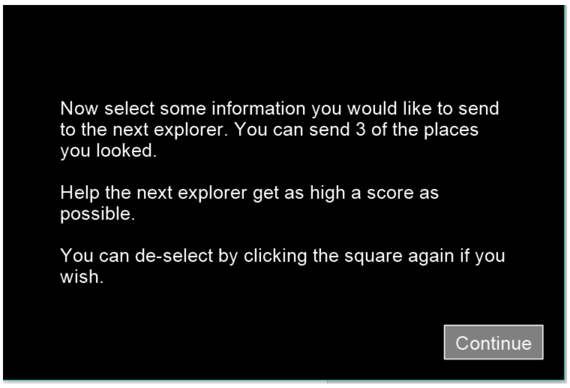


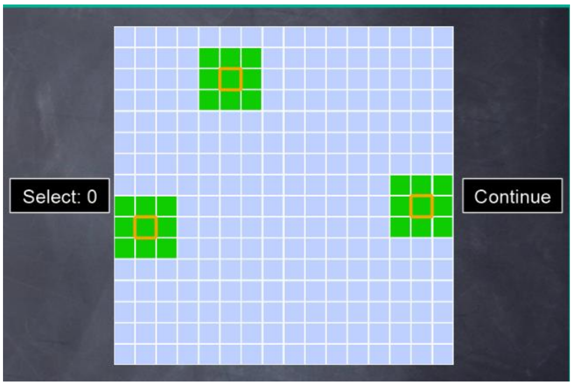

Now select some information you would like to send to the next participant. You can send 3 of the selections you made.
Your goal is to help the next participant get as high a score as possible.
You can de-select by clicking the square again if you wish.
Continue

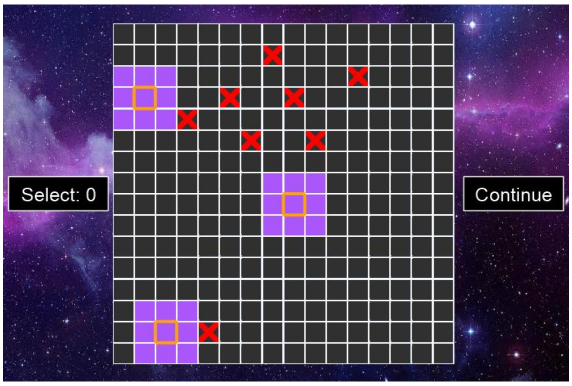

Grid 1
Continue


Lives: 9
Score: 0

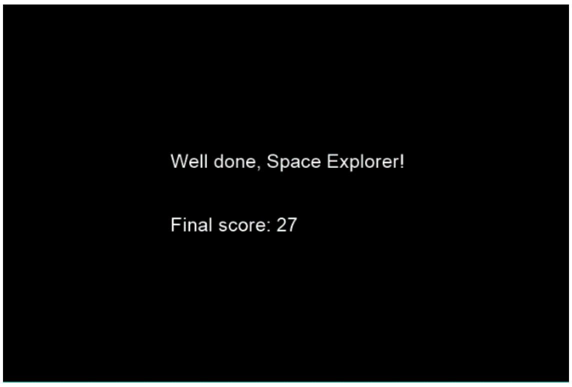

Lives: 1
Score: 27


Now select some information you would like to send to the next explorer. You can send 3 of the places you looked.
Help the next explorer get as high a score as possible.
You can de-select by clicking the square again if you wish.
Continue


Select: 0
Continue


Select: 0
Continue


That's it!
Final score: 27


Well done, Space Explorer!
Final score: 27

Appendix 8: Chapter 5 training task layout

Screen 1: Instructions

This task will give you some practise ready for the real experiment.

First, you can practice trying to sink ships in a grid. You'll get 30 lives to sink as many ships as possible, and we'll keep restocking your lives until you sink all the ships.

In the actual experiment, however, you'll only have a limited number of lives.

Screen 2: Instructions

The ships you are looking for are square, taking up 2x2 tiles. On the right of the screen you'll see a tracker of how many you've still to find.

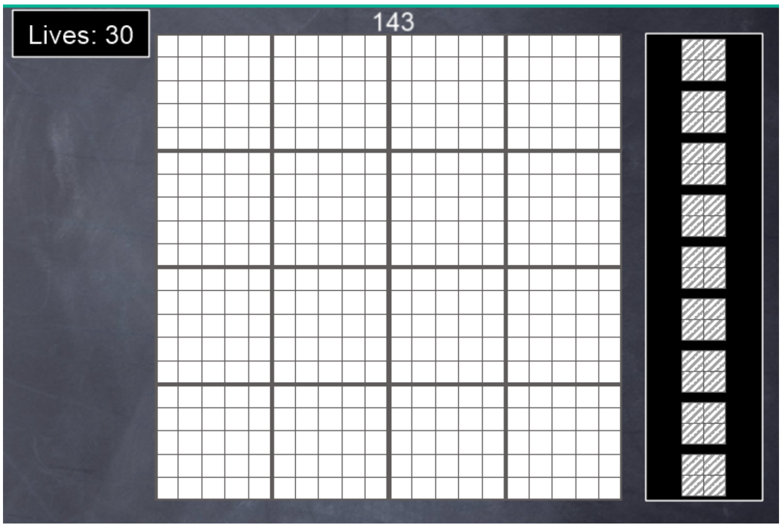
At the top of the screen you'll see how much time you have remaining to make all of your selections.

Screen 3: Instructions

Grid 1

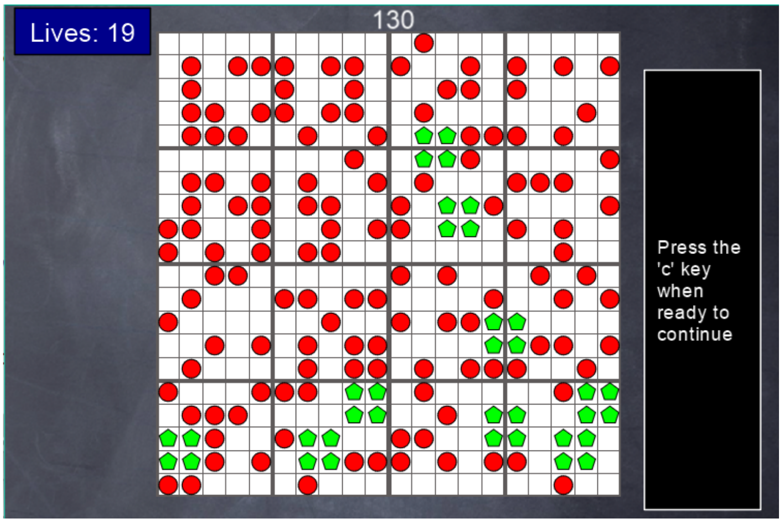
You have 30 lives to search the grid and hit as many ships as possible.

A blank training grid before the participant has made any selections.



Note. The number of search attempts remaining is shown in the top-left corner, the time remaining (number of seconds) is shown above the grid, and the number of targets left to find are shown to the right of the grid. The time remaining is an aesthetic feature to encourage participants to focus on the task only, and if the time remaining reaches 0 seconds, the task does not end, and participants can continue to make selections. Every time a participant uncovers a target, one of the targets shown to the right of the grid is highlighted and remains highlighted until the grid is complete. Participants can click on any of the grid squares to uncover whether it is a hit or a miss. Bold lines separating the grid are a visual aid only and do not imply the locations of any hidden targets.

A training grid after the participant has uncovered all of the hidden targets.



Note. Participants search the grid in sets of 30 selections, so the time remaining, and number of lives is refreshed each time. Green pentagons show uncovered hits and red circles show uncovered misses.

Screen 4: Instructions

Well done.

Now we'll let you know some more details of the experiment

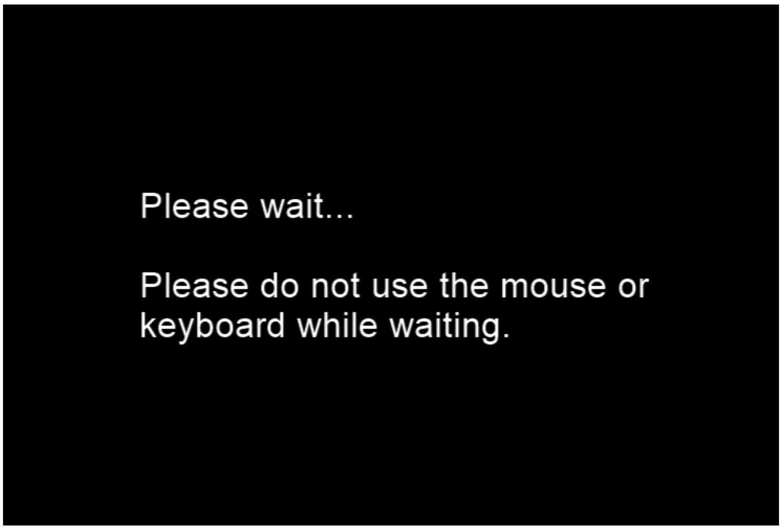
Screen 5: Instructions

You'll be doing this experiment at the same time as a partner, and often you'll have to wait for your partner to make some selections before you continue.

When this happens, you'll see a waiting screen. You might have to wait a while, but please be patient.

Press continue to see what the waiting screen looks like. It will advance automatically, so please don’t use the mouse or keyboard.

Example of the waiting screen, shown to participants



Screen 6: Instructions

At some points in the experiment, you may have the opportunity to help, or ask for help, from your partner.

All of the instructions will be given to you on screen if and when this happens, but the next screen will give you a taste of how it works.

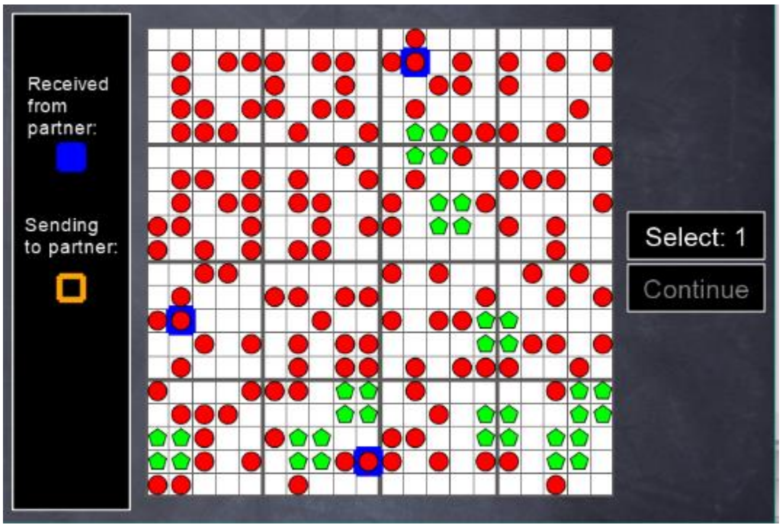
Screen 7: Instructions

Here, your partner is asking you to help them. All the huts and misses you know about will be available to you, and you must select one of those to send to your partner to try and help them to find more ships.

To help *you* do that, you may get to find out some of the tiles your partner already knows about. If you do, these will be highlighted with a blue border.

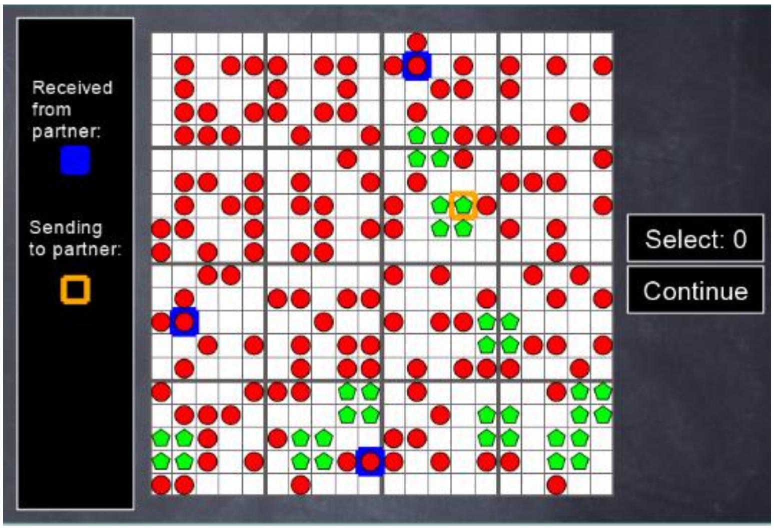
If anything isn't clear to you about this next screen, please ask one of the experimenters for more information.

A training grid during a transmission demonstration, before the current participant makes any selections.



Note. Selections highlighted with dark blue indicate selection that have been sent from a communicative partner. The number of selections remaining for the current participant is shown to the right of the grid.

A training grid during a transmission demonstration, after the current participant makes any selections.



Note. The selection made by the current participant is highlighted with orange.

Screen 8: Instructions

Well done. So in the actual experiment the selection you just made would be given to your partner to help them.

(Repeats 3 times)

Appendix 9: Full experimental task layout for Chapter 5

Left column: Brief descriptions when needed. Colour coded by communication – it fits with Figure 1. Blue is communication 1, Orange is communication 2 and Red is communication 3.

Centre column: Activity shown on PC screen 1

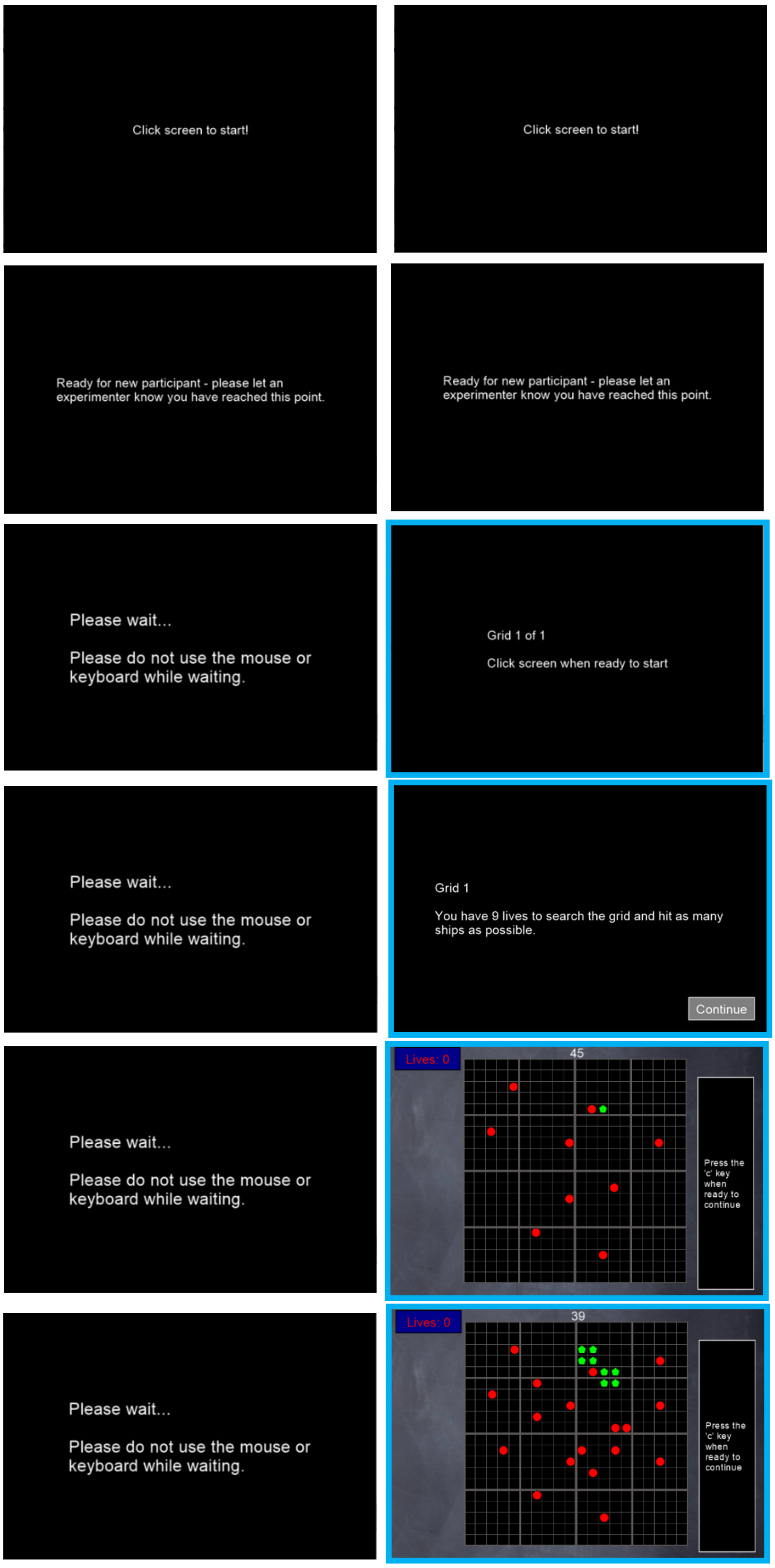
Right column: Activity shown on PC screen 2

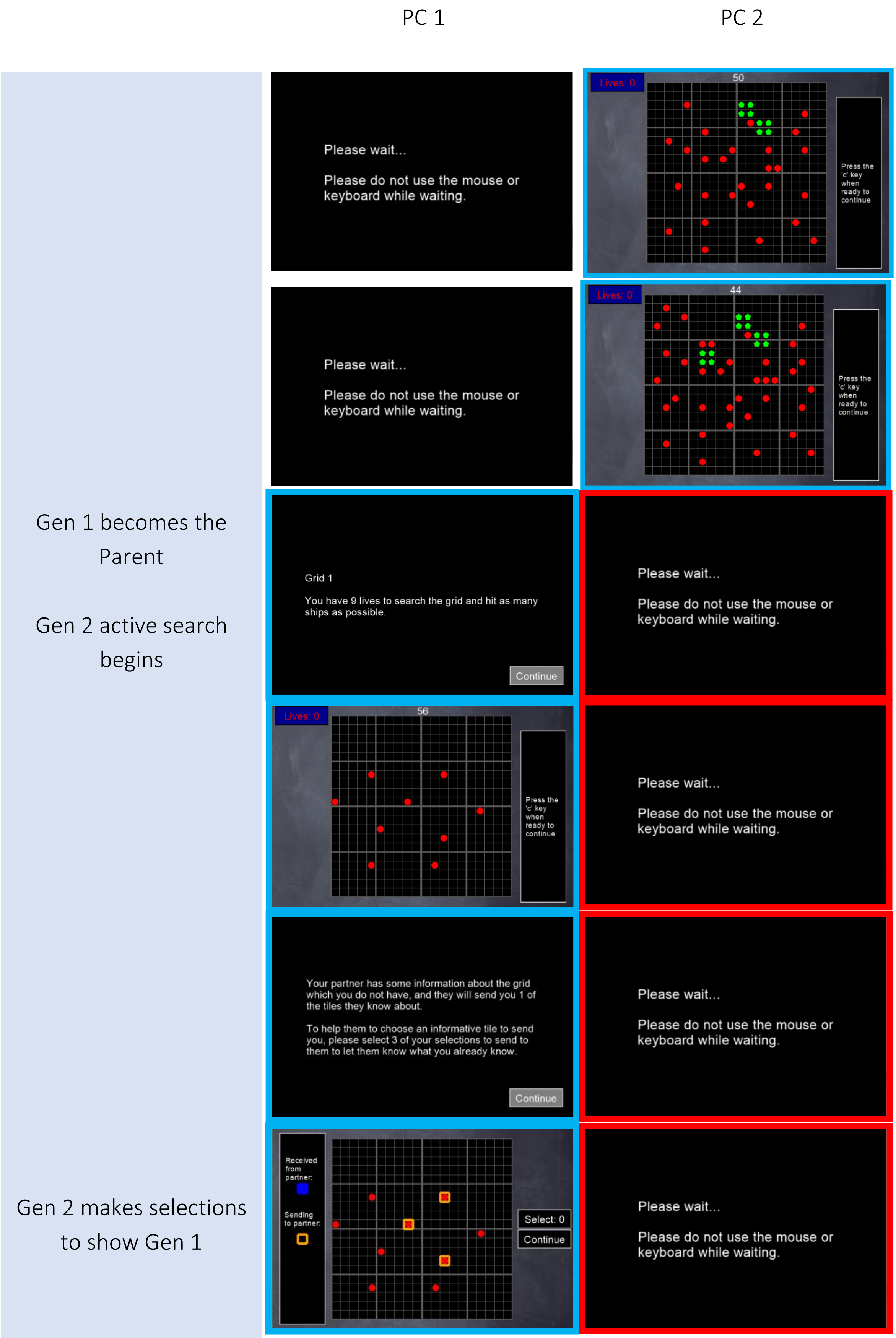
In this demonstration, participants only search one grid, however in the real task participants searched 3. Screenshots with a Red border are showing an active cultural Parent, and screenshots with a Blue border are showing an active cultural Offspring.

PC 1 PC 2

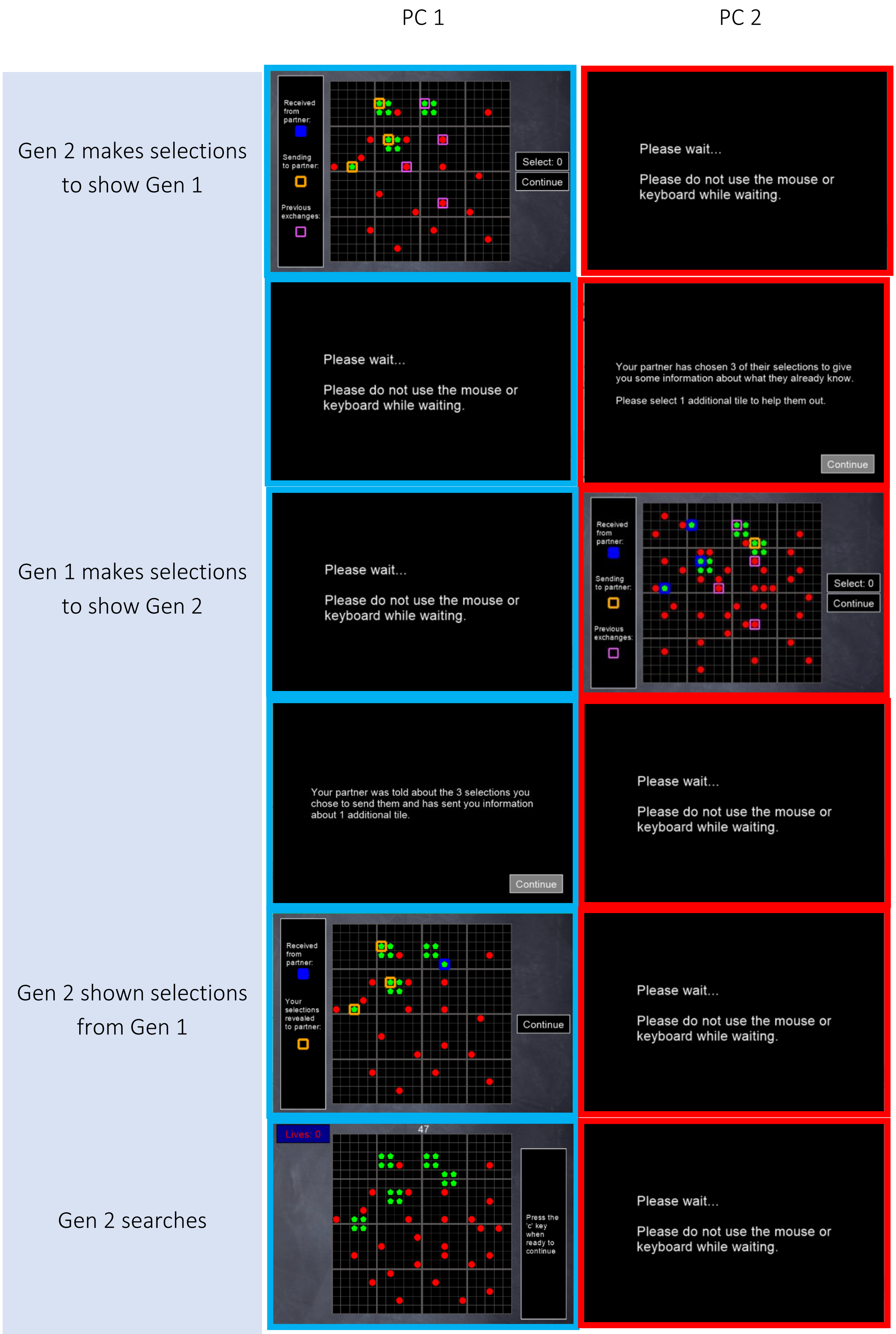
Gen 1 and Gen 2 seated

Gen 1 active search begins (searches 4 times with no communication in between)

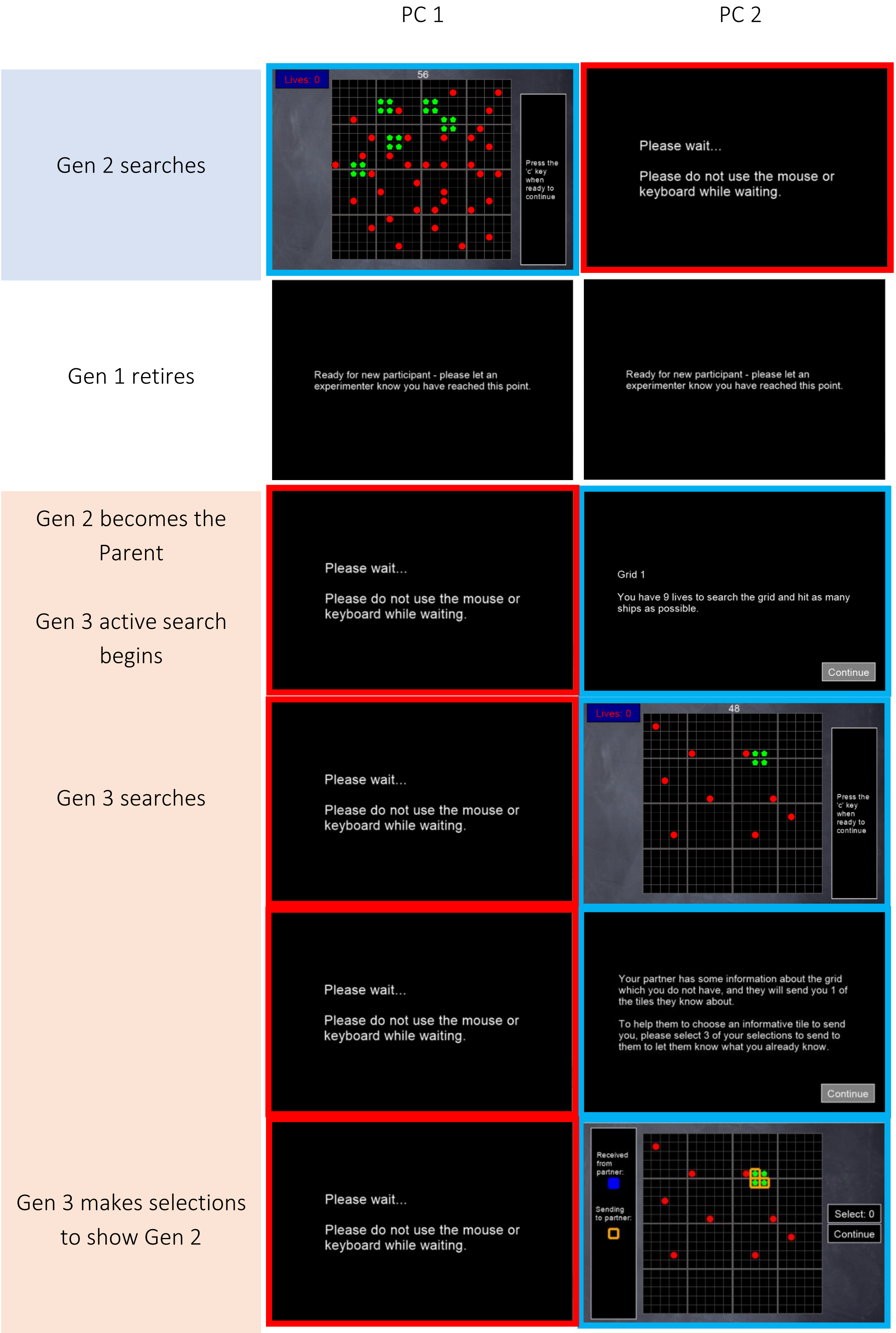



PC 1
PC 2
Please wait...
Please do not use the mouse or keyboard while waiting.
Lives: 0
50
Press the 'c' key when ready to continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Lives: 0
44
Press the 'c' key when ready to continue
Gen 1 becomes the Parent
Gen 2 active search begins
Grid 1
You have 9 lives to search the grid and hit as many ships as possible.
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Lives: 0
56
Press the 'c' key when ready to continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Your partner has some information about the grid which you do not have, and they will send you 1 of the tiles they know about.
To help them to choose an informative tile to send you, please select 3 of your selections to send to them to let them know what you already know.
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Gen 2 makes selections to show Gen 1
Received from partner:
Sending to partner:
Select: 0
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.

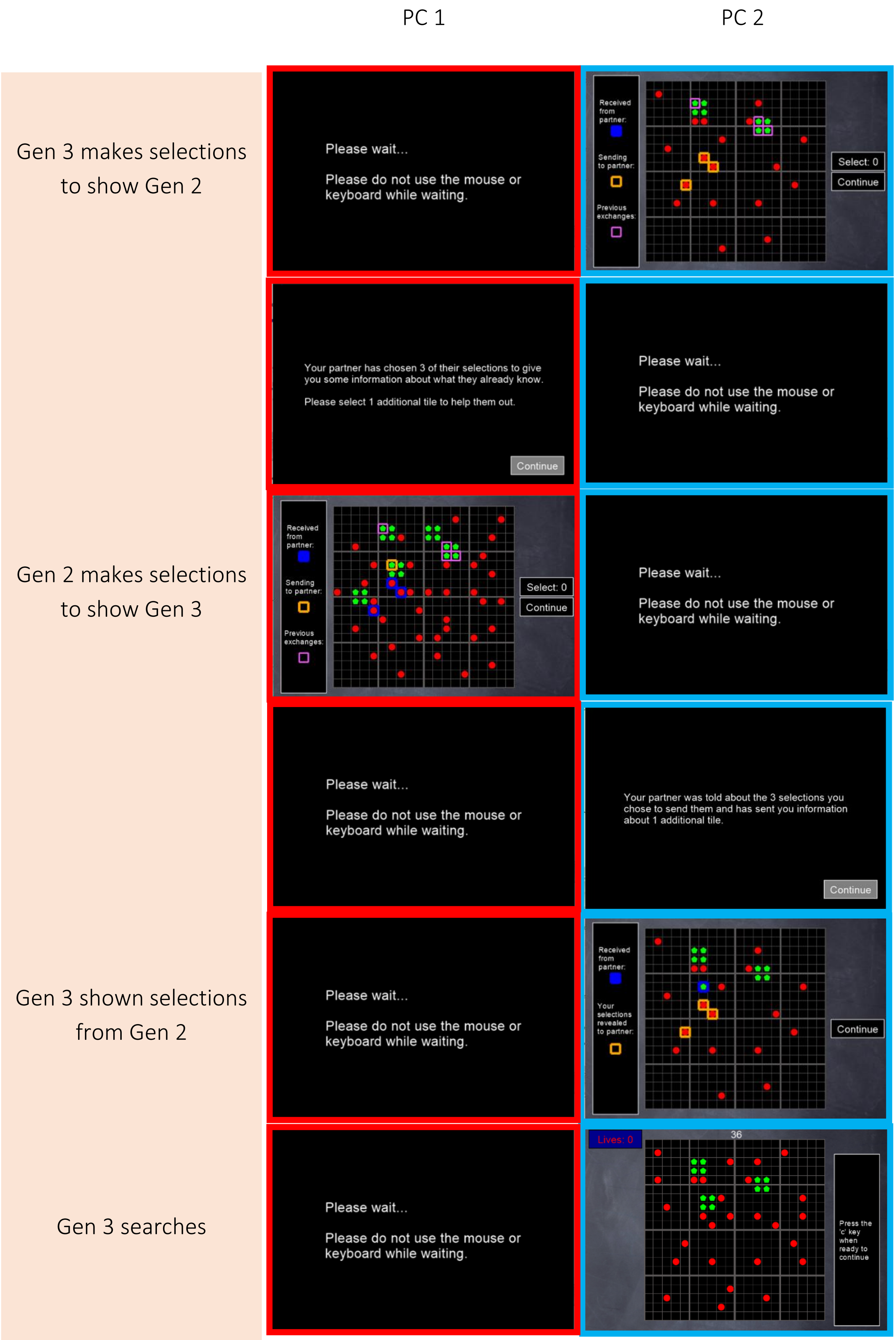
PC 1
PC 2
Gen 1 makes selections to show Gen 2
Please wait…
Please do not use the mouse or keyboard while waiting.
Your partner has chosen 3 of their selections to give you some information about what they already know.
Please select 1 additional tile to help them out.
Continue
Please wait…
Please do not use the mouse or keyboard while waiting.
Received from partner:
Sending to partner:
Select: 0
Continue
Your partner was told about the 3 selections you chose to send them and has sent you information about 1 additional tile.
Continue
Please wait…
Please do not use the mouse or keyboard while waiting.
Gen 2 shown selections from Gen 1
Received from partner:
Your selections revealed to partner:
Continue
Please wait…
Please do not use the mouse or keyboard while waiting.
Gen 2 searches
Lives: 0
40
Press the 'c' key when ready to continue
Please wait…
Please do not use the mouse or keyboard while waiting.
Your partner has some information about the grid which you do not have, and they will send you 1 of the tiles they know about.
To help them to choose an informative tile to send you, please select 3 of your selections to send to them to let them know what you already know.
Continue
Please wait…
Please do not use the mouse or keyboard while waiting.


PC 1
PC 2
Gen 2 makes selections to show Gen 1
Received from partner:
Sending to partner:
Previous exchanges:
Select: 0
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Please wait...
Please do not use the mouse or keyboard while waiting.
Your partner has chosen 3 of their selections to give you some information about what they already know.
Please select 1 additional tile to help them out.
Continue
Gen 1 makes selections to show Gen 2
Please wait...
Please do not use the mouse or keyboard while waiting.
Received from partner:
Sending to partner:
Previous exchanges:
Select: 0
Continue
Your partner was told about the 3 selections you chose to send them and has sent you information about 1 additional tile.
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Gen 2 shown selections from Gen 1
Received from partner:
Your selections revealed to partner:
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Gen 2 searches
Lives: 0
47
Press the 'c' key when ready to continue
Please wait...
Please do not use the mouse or keyboard while waiting.

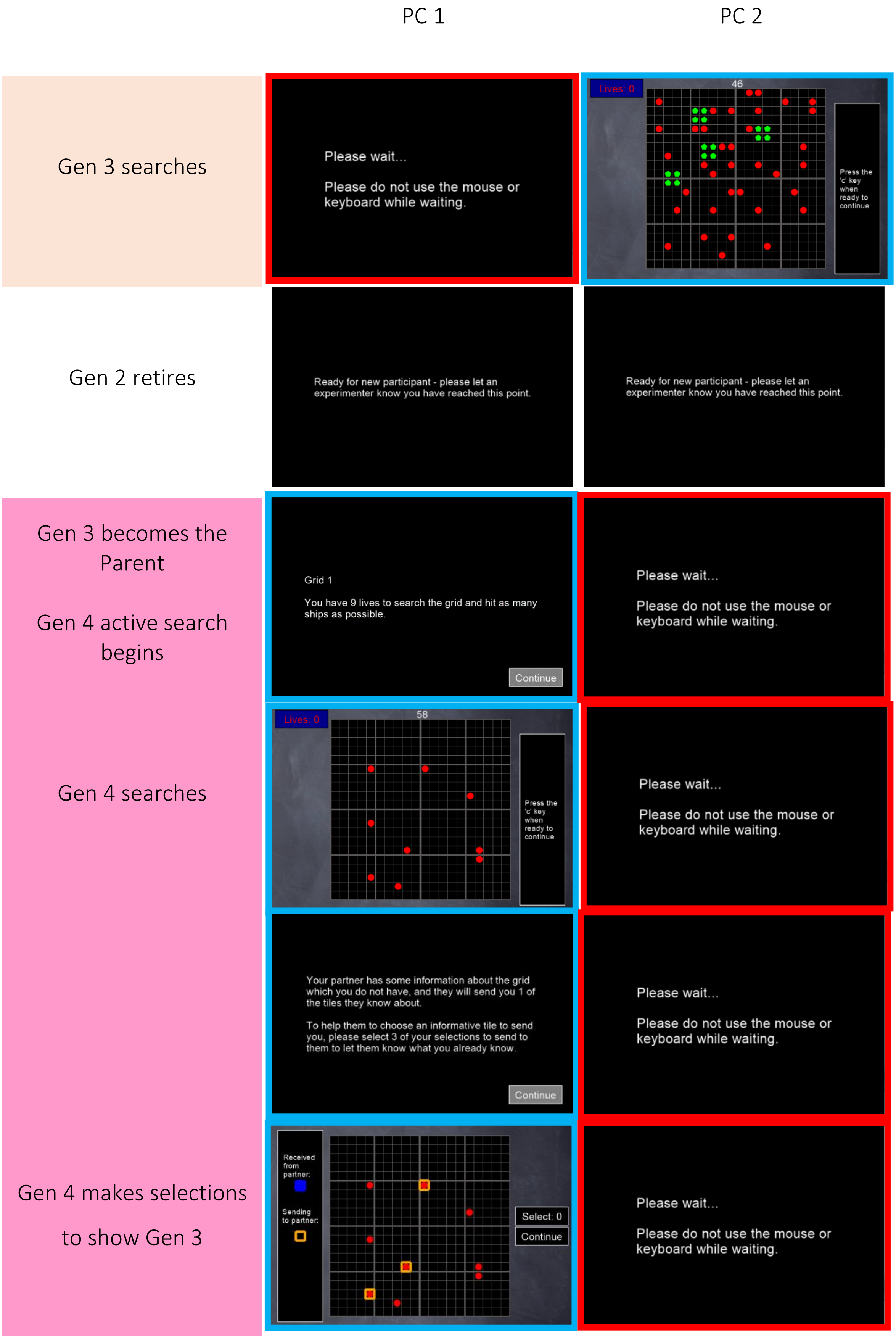

PC 1
PC 2
Gen 2 makes selections to show Gen 1
Gen 1 makes selections to show Gen 2
Gen 2 shown selections from Gen 1
Your partner has some information about the grid which you do not have, and they will send you 1 of the tiles they know about.
To help them to choose an informative tile to send you, please select 3 of your selections to send to them to let them know what you already know.
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Received from partner:
Sending to partner:
Previous exchanges:
Select: 0
Continue
Your partner has chosen 3 of their selections to give you some information about what they already know.
Please select 1 additional tile to help them out.
Your partner was told about the 3 selections you chose to send them and has sent you information about 1 additional tile.
Your selections revealed to partner:


PC 1
PC 2
Gen 2 searches
Lives: 0
56
Press the 'c' key when ready to continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Gen 1 retires
Ready for new participant - please let an experimenter know you have reached this point.
Ready for new participant - please let an experimenter know you have reached this point.
Gen 2 becomes the Parent
Gen 3 active search begins
Please wait...
Please do not use the mouse or keyboard while waiting.
Grid 1
You have 9 lives to search the grid and hit as many ships as possible.
Continue
Gen 3 searches
Please wait...
Please do not use the mouse or keyboard while waiting.
Lives: 0
48
Press the 'c' key when ready to continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Your partner has some information about the grid which you do not have, and they will send you 1 of the tiles they know about.
To help them to choose an informative tile to send you, please select 3 of your selections to send to them to let them know what you already know.
Continue
Gen 3 makes selections to show Gen 2
Please wait...
Please do not use the mouse or keyboard while waiting.
Received from partner:
Sending to partner:
Select: 0
Continue


PC 1
PC 2
Gen 2 makes selections to show Gen 3
Gen 3 shown selections from Gen 2
Gen 3 searches
Your partner has chosen 3 of their selections to give you some information about what they already know.
Please select 1 additional tile to help them out.
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Received from partner:
Sending to partner:
Select: 0
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Please wait...
Please do not use the mouse or keyboard while waiting.
Your partner was told about the 3 selections you chose to send them and has sent you information about 1 additional tile.
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Received from partner:
Your selections revealed to partner:
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Lives: 0
48
Press the 'c' key when ready to continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Your partner has some information about the grid which you do not have, and they will send you 1 of the tiles they know about.
To help them to choose an informative tile to send you, please select 3 of your selections to send to them to let them know what you already know.
Continue


PC 1
PC 2
Gen 3 makes selections to show Gen 2
Please wait...
Please do not use the mouse or keyboard while waiting.
Received from partner:
Sending to partner:
Previous exchanges:
Select: 0
Continue
Your partner has chosen 3 of their selections to give you some information about what they already know.
Please select 1 additional tile to help them out.
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Gen 2 makes selections to show Gen 3
Received from partner:
Sending to partner:
Previous exchanges:
Select: 0
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Please wait...
Please do not use the mouse or keyboard while waiting.
Your partner was told about the 3 selections you chose to send them and has sent you information about 1 additional tile.
Continue
Gen 3 shown selections from Gen 2
Please wait...
Please do not use the mouse or keyboard while waiting.
Received from partner:
Your selections revealed to partner:
Continue
Gen 3 searches
Please wait...
Please do not use the mouse or keyboard while waiting.
Lives: 0
36
Press the 'c' key when ready to continue

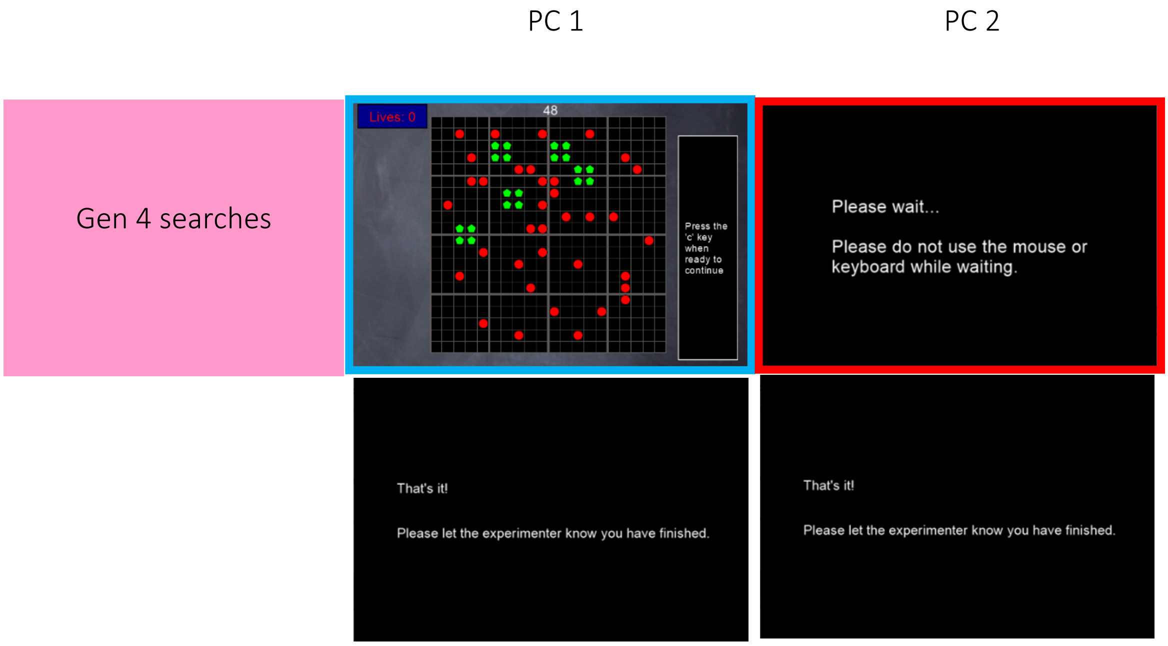
PC 1
PC 2
Gen 3 makes selections to show Gen 2
Please wait...
Please do not use the mouse or keyboard while waiting.
Your partner has some information about the grid which you do not have, and they will send you 1 of the tiles they know about.
To help them to choose an informative tile to send you, please select 3 of your selections to send to them to let them know what you already know.
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Received from partner:
Sending to partner:
Previous exchanges:
Select: 0
Continue
Gen 2 makes selections to show Gen 3
Your partner has chosen 3 of their selections to give you some information about what they already know.
Please select 1 additional tile to help them out.
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Received from partner:
Sending to partner:
Previous exchanges:
Select: 0
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Please wait...
Please do not use the mouse or keyboard while waiting.
Your partner was told about the 3 selections you chose to send them and has sent you information about 1 additional tile.
Continue
Gen 3 shown selections from Gen 2
Please wait...
Please do not use the mouse or keyboard while waiting.
Received from partner:
Your selections revealed to partner:
Continue


PC 1
PC 2
Gen 3 searches
Please wait...
Please do not use the mouse or keyboard while waiting.
Lives: 0
46
Press the 'c' key when ready to continue
Gen 2 retires
Ready for new participant - please let an experimenter know you have reached this point.
Ready for new participant - please let an experimenter know you have reached this point.
Gen 3 becomes the Parent
Gen 4 active search begins
Grid 1
You have 9 lives to search the grid and hit as many ships as possible.
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Gen 4 searches
Lives: 0
58
Press the 'c' key when ready to continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Your partner has some information about the grid which you do not have, and they will send you 1 of the tiles they know about.
To help them to choose an informative tile to send you, please select 3 of your selections to send to them to let them know what you already know.
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Gen 4 makes selections to show Gen 3
Received from partner:
Sending to partner:
Select: 0
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.

PC 1
PC 2
Gen 3 makes selections to show Gen 4
Please wait...
Please do not use the mouse or keyboard while waiting.
Your partner has chosen 3 of their selections to give you some information about what they already know.
Please select 1 additional tile to help them out.
Continue
Received from partner:
Sending to partner:
Select: 0
Continue
Your partner was told about the 3 selections you chose to send them and has sent you information about 1 additional tile.
Continue
Gen 4 shown selections from Gen 3
Received from partner:
Your selections revealed to partner:
Continue
Gen 4 searches
Lives: 0
48
Press the 'c' key when ready to continue
Your partner has some information about the grid which you do not have, and they will send you 1 of the tiles they know about.
To help them to choose an informative tile to send you, please select 3 of your selections to send to them to let them know what you already know.
Continue


PC 1
PC 2
Gen 4 makes selections to show Gen 3
Received from partner:
Sending to partner:
Previous exchanges:
Select: 0
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Your partner has chosen 3 of their selections to give you some information about what they already know.
Please select 1 additional tile to help them out.
Continue
Gen 3 makes selections to show Gen 4
Please wait...
Please do not use the mouse or keyboard while waiting.
Received from partner:
Sending to partner:
Previous exchanges:
Select: 0
Continue
Your partner was told about the 3 selections you chose to send them and has sent you information about 1 additional tile.
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Gen 4 shown selections from Gen 3
Received from partner:
Your selections revealed to partner:
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Gen 4 searches
Lives: 0
53
Press the 'c' key when ready to continue
Please wait...
Please do not use the mouse or keyboard while waiting.


PC 1
PC 2
Gen 4 makes selections to show Gen 3
Gen 3 makes selections to show Gen 4
Gen 4 shown selections from Gen 3
Your partner has some information about the grid which you do not have, and they will send you 1 of the tiles they know about.
To help them to choose an informative tile to send you, please select 3 of your selections to send to them to let them know what you already know.
Continue
Please wait...
Please do not use the mouse or keyboard while waiting.
Received from partner:
Sending to partner:
Previous exchanges:
Select: 0
Continue
Your partner has chosen 3 of their selections to give you some information about what they already know.
Please select 1 additional tile to help them out.
Your partner was told about the 3 selections you chose to send them and has sent you information about 1 additional tile.
Your selections revealed to partner:


PC 1
PC 2
Gen 4 searches
Lives: 0
48
Press the 'c' key when ready to continue
Please wait...
Please do not use the mouse or keyboard while waiting.
That's it!
Please let the experimenter know you have finished.
That's it!
Please let the experimenter know you have finished.