ON THE DEVELOPMENTAL ORIGINS OF HUMAN MATERIAL CULTURE

by

EVA MARIA REINDL

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

School of Psychology

College of Life and Environmental Sciences

University of Birmingham

June 2017

Abstract

Material culture – tools, technology, and instrumental skills – has allowed humans to live in almost every habitat on earth. This thesis investigates the developmental roots of human material culture by examining basic tool-use skills and cultural learning abilities in young children. The introduction presents the concepts of the Zone of Latent Solutions (Tennie, Call, & Tomasello, 2009), cumulative culture, and Vygotsky's (1978) theories as the theoretical background for the following five experiments. Chapter 2 identifies a list of tool-use behaviours that children can invent *individually* and thus represent an ontogenetic and phylogenetic basis of human tool culture. Chapter 3 extends this list by several behaviours involving the use of two tools in combination (Associative tool use). Chapter 4 focuses on a cultural behaviour that children can only acquire *socially*. It uses an adapted version of the spaghetti tower task (Caldwell & Millen, 2008a) to study whether children can copy a material cultural product that they could not have invented on their own and whether they can do so without action information. Chapter 5 uses the same task to investigate whether groups of children can produce a ratchet effect. The discussion summarizes the findings and presents limitations and directions for future research.

Acknowledgements

A PhD is a journey that is never travelled alone. I am deeply thankful to everyone who supported me on my way and made the last four years such an enriching time.

My first thanks go to my brilliant supervisors who welcomed me so warmly in Birmingham and who placed great confidence in me from the start – I could always count on them. I thank Claudio Tennie for his continuous encouragement to think big and aim high, to work independently, and for his generous financial support without which I would not have been able to attend so many enriching conferences and workshops. I thank Sarah Beck for the innumerable big and small pieces of advice, for always having a friendly ear, for the continuous support, and all the enthusiasm brought into my project. I thank Ian Apperly for his invaluable intellectual input and for challenging and promoting my thinking – he is a great role model.

I thank the College of Life and Environmental Sciences at the University of Birmingham for granting me a PhD studentship without which I would not have been able to conduct my research nor to attend any conferences. Financial support also came from the School of Psychology who funded my travel and accommodation when I was attending a statistics course at the Max-Planck-Institute for Evolutionary Anthropology in Leipzig, Germany.

I thank all the participants in my studies, especially the children and their parents (we did test adults, too). It was a pleasure working with the kids, and we had some pretty good laughs together. I thank all the schools and nurseries in the UK and in Germany who allowed me to carry out my research. I am especially thankful for the support I received from Ms

Birgit Frank from the Kindergarten Rötz, Germany. Special thanks go to Thinktank Science Museum and to Lynsey Rutter and Laurence Butler who have allowed me so many times to carry out my studies at the museum and who provided their resources and facilities to support my research. All this would not have been possible without this collaboration. I also thank Twycross Zoo and Manuela Townsend for establishing a collaboration with us, letting me conduct one of my studies at their facilities, and letting me stay at their accommodation.

I thank my friends and colleagues at Birmingham. Thank you to Elisa Bandini for being a great friend and an irreplaceable lab buddy whom I can always count on. I think we really created something within the last three years. I thank Damien Neadle for very helpful discussions and comments on the theories presented in this thesis. My thanks also go to my housemate Bram Geron, for all the great discussions and for listening. Special thanks go to Nuredin Ibrahim Mohammed – for everything.

I am deeply thankful for the unconditional love and support I have received for 28 years now by my parents Hans and Maria Reindl who have never hesitated to help me out in any way. I have never seen any parents who took it for granted to drive 3000 km for dropping their child’s stuff into a new home. Thank you, Pa, for your infinite kindness, patience, and of course for helping me build my apparatuses. Thank you, Mam, for your selfless support, for always being there, and for being the best mom ever.

Lastly, from the bottom of my heart, my biggest thanks go to my sister Hanna, who has not just supported my journey but who fully embarked on it with me. Saying thank you is not enough to acknowledge how much you had to put back so that I could go on this journey. I am so grateful for your unconditional love, your patience, for saving me from so many things. I love you.

Table of Contents

List of Illustrations

List of Tables

Abbreviations

ANCOVA = Analysis of Covariance

ATU = Associative Tool Use

CC = Cumulative Culture

CI = Confidence Interval

DV = Dependent Variable

E = Experimenter

GATTeB = Great Ape Tool Test Battery

GLM = Generalized Linear Model

GLMM = Generalized Linear Mixed Model

IV = Independent Variable

LMM = Linear Mixed Model

OR = Odds Ratio

SD = Standard Deviation

SE = Standard Error

VIF = Variance Inflation Factor

ZAD = Zone of Actual Development

ZLS = Zone of Latent Solutions

ZPD = Zone of Proximal Development

CHAPTER 1: INTRODUCTION

1.1 General introduction

"The key to understanding how humans evolved and why we are so different from other animals is to recognize that we are a cultural species." (Henrich, 2015, p. 3)

For many of us, *Homo sapiens* is among the most fascinating and extraordinary species we know. Humans have come to live in almost every terrestrial environment on earth – an achievement not equaled by any other vertebrate (Boyd, Richerson, & Henrich, 2011; Henrich, 2015) – and we are the only species reaching out to habitats beyond our planet (NASA, 2016; Smith, 2016). Not least, we are the only species trying to understand our kind and why we appear so different from others.

An increasing number of researchers argue that the key to explaining humans' uniqueness is acknowledging that we are "a cultural species" (Henrich, 2015, p. 3, see introductory quote; see also Laland, Odling-Smee, & Feldman, 2000; Richerson & Boyd, 2005; Tomasello, 1999a, 1999b; but see Cosmides & Tooby 2001; Pinker, 2010). That is, these researchers argue that the "secret of our success" (Henrich, 2015) is not so much the fact that individual humans are more intelligent than other animals, but that it is our ability to share, transmit, and accumulate knowledge that equips us with the necessary means to adapt to diverse environments:

> To deal with everything from the Arctic to the tropics, humans as a species have evolved a highly flexible suite of cognitive skills. But these are not individual cognitive skills that enable individuals to survive alone in the tundra or rain forest, but rather they are social-cognitive skills that enable them to develop, in concert with others in their cultural groups, creative ways of coping with whatever challenges may arise. (Tomasello & Herrmann, 2010, p. 7)

In other words, we are smart not because *individual* brains possess exceptional capacities, but because our ability to learn from others faithfully and build upon their knowledge creates a form of *collective* intelligence which is stored in the heads of many members of our societies (*collective brain hypothesis,* Muthukrishna & Henrich, 2016; see also Herrmann, Call, Hernández-Lloreda, Hare, & Tomasello, 2007; Tomasello & Moll, 2010). Over evolutionary time, our (social and asocial) learning abilities on the one hand and the range and efficiency of our cultural repertoire on the other hand have coevolved: Culture has made our species smart (*phylogenetic cultural intelligence hypothesis*, Tennie & Over, 2012; van Schaik & Pradhan, 2003). A similar logic applies to the level of the individual: Cultural learning over the life course makes individuals smart. Growing up in a group and acquiring its knowledge, skills, language, and customs, and interacting with its cultural artefacts equip children with novel mental tools. These mental tools might include the use of signs and memory techniques (Vygotsky, 1978), analogies (Tomasello, 1999a), or the concept of zero (Henrich, 2015) which can qualitatively change the possessors' way of thinking (*ontogenetic cultural intelligence hypothesis*, Herrmann et al., 2007; Tennie & Over, 2012; van Schaik & Burkart, 2011). Not only does culture provide us with useful upgrades to our cognitive machinery, it is in fact necessary for the full development of our species-typical cognitive abilities – and likely even for our survival (see e.g., the essential role of clothes and cooking for our survival, Henrich, 2015; Tomasello, 1999a).

The evolution of our culture has taken place in the domains of *material* and *social conventional* culture (Legare & Nielsen, 2015). Material culture describes the tools, artefacts, technology, instrumental skills and knowledge of a given population. Social conventional culture describes the "group-specific practices that function to increase cohesion and cooperation among group members" (ibid., p. 1) – e.g., rituals or traffic laws – where these practices may or may not involve objects. The two types of culture are not strictly separated, but best understood as two extremes of a continuum (ibid.). This thesis focuses on the domain of material culture. Material culture supports individuals in their interaction with the environment and with themselves (e.g., when using a stick to scratch oneself). Compared to social conventional culture, material culture tends to be directly "oriented towards solving concrete environmental problems" (Boesch, 2012, p. 79), allowing individuals to shape their environment, relax some of its restrictions imposed on them, and to open up new ecological niches (Boesch, 2003, 2012). Material culture is an especially powerful form of *niche construction*, describing the idea that individuals – by acting on and interacting with their environment – alter their environment, with these alterations generating feedback and new selection pressures for the individual and being "inherited" by the next generation (Laland & O'Brien, 2012; Laland et al., 2000).

Material culture has played a significant role in human evolution (Coolidge & Wynn, 2005; Gergely & Csibra, 2006; Stout, Toth, & Schick, 2000): Over evolutionary time, our ancestors increasingly relied on tools and technology to solve problems, survive, and spread, leading to the coevolution of our material culture (i.e., its size and complexity) and our (socio-) cognitive capacities such as social and asocial learning, working memory, creativity, inventive abilities, and cooperation, eventually producing the human "technological niche" (Stout & Khreisheh, 2015, p. 867) – a package of tools, skills, and knowledge that support our survival but are too complex for individuals to invent on their own (see also Ambrose, 2001; Richerson & Boyd, 2005; Coolidge & Wynn, 2005; Sterelny, 2012). And so, our ability to produce the current forms of our material culture is both explanandum and explanans in the study of humans' uniqueness.

Research on the origins of human (material) culture is inherently multidisciplinary (Mesoudi, 2011). As a developmental psychologist I aimed to contribute with this thesis some new insights into the ontogenetic origins of human material culture by carrying out behavioural experiments with children. In recent years, there has been a strong focus on children's social learning to investigate how children acquire novel cultural skills, e.g., tool-using behaviours (see below). A hundred years ago, Vygotsky (1978) even believed that all tool use in humans had to be learned socially. Not much was known about which tool-use behaviours young children can invent on their own, i.e., asocially. However, research on this ontogenetic basis of human tool culture is needed: Knowing which tool behaviours children can invent on their own helps us identify those behaviours that actually rely on social learning to be acquired and allows us to describe the ontogenetic behavioural basis for socially learned tool behaviours. Therefore, the first part of this thesis examined the types of tool-use behaviours young children can invent on their own, without social learning (chapters 2 and 3). The second half of the thesis moves away from children's *spontaneous* tool-use abilities, focusing on children's capacity to *socially learn* from, copy, and improve material culture (chapters 4 and 5).

During the last 100 years three theoretical concepts have been developed which have proven to be a useful, cross-disciplinary framework for describing, investigating, and comparing the cultural abilities of human and non-human animals: the *zone of latent solutions* theory (Tennie, Call, & Tomasello, 2009), the notion of *cumulative culture* (Boyd & Richerson, 1996), and Vygotsky's (1978) *zone of proximal development*. In the next section I introduce these concepts as the theoretical background of this thesis.

1.2 Theoretical and empirical background

> *"No animal comes close to having humans' ability to build on previous discoveries and pass the improvements on. What determines those differences could help us understand how human culture evolved."* (Galef, 2009, p. 242)

1.2.1 The phylogenetic roots of (human) culture: The zone of latent solutions (ZLS)

1.2.1.1 What is the ZLS?

Social learning, i.e., "learning that is facilitated by observation of, or interaction with, another individual or its products" (Hoppitt & Laland, 2013, p. 4) is widespread in the animal kingdom (Laland & Hoppitt, 2003). However, the ability to both socially learn *and* further improve learned skills beyond a point achievable by any individual – a phenomenon that has been called *the ratchet effect* (Tomasello, Kruger, & Ratner, 1993) – seems to be unique to human groups (Dean, Vale, Laland, Flynn, & Kendal, 2014; but see Sasaki & Biro, 2017). In order to explain why humans exhibit the ratchet effect whereas such a capacity seems to be lacking in our closest living relatives – the great apes – Tennie et al. (2009) put forward the zone of latent solutions (ZLS) theory (see also Reindl, Bandini, & Tennie, in press; Tennie & Hedwig, 2009). The theory aims to identify differences and similarities between cultures of different species in order to contribute to our understanding of how human culture may have evolved (see introductory quote to this section). The definition of *culture* used in this thesis will follow the definition of Whiten and van Schaik (2007) as the "possession of multiple traditions, spanning different domains of behaviour" (p. 605), with *tradition* describing a "distinctive behaviour pattern shared by two or more individuals in a social unit, which persists over time and that new practitioners acquire in part through socially aided learning" (Fragaszy & Perry, 2003, p. xiii).

The ZLS theory suggests that many – perhaps all – animals possess a range of behavioural traits that can potentially be (re)invented by any able-bodied member of the species, without the need for social learning (*latent solutions*). This does not mean that social learning plays no role in the likelihood of expression of latent solutions. Social learning can put the learner into a favourable situation in which it will be more likely to individually (re)invent the behaviour than when no social learning opportunities were available. For example, the learner can be exposed to relevant materials and environments just because it stays close to its group (*exposure*, Tomasello, 1999a) or the activity of others can draw the individual's attention towards a certain location (*local enhancement*, Thorpe, 1956) or stimulus (*stimulus enhancement*, Spence, 1937). With the help of social learning mechanisms such as these, the individual is more likely to develop the behaviour itself by ultimately engaging in asocial (including trial-and-error) learning. The resulting acquired behaviour is therefore an *individual reinvention* (a latent solution) rather than a *social copy* (Reindl et al., in press). The role of social learning would then be restricted to affecting the likelihood of an individual's invention of a latent solution. Across individuals, the sum of such effects can lead to a homogeneous expression of the behaviour within a group (i.e., almost every group member showing the behaviour), outwardly resembling a spread through social learning – however, social learning is not actually necessary for the reinvention of the behaviour[1].

Given various findings and reviews in the literature, the ZLS theory asserts that the potentially cultural (e.g., tool-use) behaviours of great apes all represent latent solutions. Wild chimpanzees (Boesch, 2012; Langergraber et al., 2011; Whiten et al., 1999, 2001) and orangutans (van Schaik et al., 2003, 2009) exhibit a range of tool-use behaviours for food acquisition, personal hygiene, communicative purposes, etc., such as using sticks to fish for insects, leaves to wipe the body, or slapping a branch to attract someone's attention. These behaviours show substantial variation in their frequency across the communities of the respective species. Some behaviours are present in many communities (e.g., using sticks to dip for fluids in chimpanzees, Whiten et al., 2001), others occur only rarely (e.g. using sticks to scoop for algae in chimpanzees, ibid.). In their reviews of these potentially cultural behaviours researchers have used the *method of elimination* (van Schaik, 2003), i.e., they applied systematic criteria to rule out the possibility that these differences in distributions can solely be explained by ecological or genetic factors, in order to find out which of these behaviours are potentially cultural, i.e., influenced by social learning (Langergraber et al., 2011; van Schaik et al., 2003, 2009; Whiten et al.,

[1] Note that the claim that latent solutions do not require social learning in order to be acquired does not rule out the possibility that in species capable of faithful social learning, i.e., of closely matching not only the results of an observed action, but also the precise means and the mechanics of the action (e.g., as humans often do), there could be cases in which learners actually acquire a latent solution through imitation (even though they would not need to revert to social learning).

1999, 2011). For example, only those behaviours that were found to be customary (i.e., present in almost all “able-bodied members of at least one age-sex class”, Whiten et al., 1999, p. 682) or habitual (i.e., not customary, but repeatedly shown by several individuals) in some communities but absent in others (where this absence could not be explained by ecological reasons) were suggested as being potentially cultural. Behaviours that resembled this pattern (customary or habitual in some communities but absent in others) but for which the absence could be explained by ecological factors were excluded from the list of potentially cultural behaviours. For example, chimpanzees’ use of a rock to prop a stone anvil for nutcracking was not counted as a cultural behaviour as those chimpanzee communities lacking anvil propping did not have suitable rocks in the first place.

Subsequent studies further investigated potential effects of ecological factors on tool behaviours, e.g., how specific ant prey species affect chimpanzees’ ant eating behaviours (Möbius, Boesch, Koops, Matsuzawa, & Humle, 2008; Schöning, Humle, Möbius, & McGrew, 2008). These studies also suggested that while some behavioural differences between sites (e.g., differences in tool length and technique) could be attributed to differences in ecological factors (e.g., the aggressiveness of the ant species), the geographic distribution of ant eating behavioural variants could not fully be explained by ecological factors, thus suggesting the existence of cultural differences. In addition, studies comparing genetically similar neighbouring chimpanzee communities (“eliminating” the factor “genetics”) that live in the same environment (“eliminating” differences in the environment) have also suggested cultural differences between sites, e.g., in nutcracking behaviours (Luncz, Mundry, & Boesch, 2012; van Leeuwen, Cronin, Haun, Mundry, & Bodamer, 2012).

According to the ZLS theory, the great ape behaviours identified as cultural can be reinvented by any able-bodied individual of the respective species on its own. Whether or not a given behaviour is actually expressed depends on the presence of a favourable constellation of social and ecological factors (Tennie et al., 2009). Once invented, social learning opportunities and social learning biases (e.g., biases with regard to whom or under which circumstances to copy, Hoppitt & Laland, 2013) can influence whether and how many other group members also invent the behaviour. In this way, each great ape community may have come to possess their unique profile of cultural behaviours (or realised latent solutions; e.g., population A might show behaviour X and Z, while population B might show behaviour Y and Z), superficially resembling the case of human cultures (see Fig. 1 in Whiten et al., 1999, and Whiten, 2005). Thus, the ZLS theory treats these great ape behaviours as potentially cultural not in the sense that the behaviours spread through social learning but in the sense that these behaviours are individual reinventions with the likelihood of these reinventions being influenced by social learning.

The hypothesis that wild great apes' tool-use behaviours are latent solutions, i.e., do not rely on social learning to be acquired, can be tested via *latent solution tests.* These tests present captive, target-behaviour naïve individuals with all the raw material required to invent the tool behaviour in question. If it is found that two or more independent members of a species spontaneously invent the behaviour, the behaviour is assumed to fall within the species' ZLS (for further explanation of the "two or more individuals" approach, see chapter 2)[2]. There are already a few latent solution tests with great apes (as well as with

[2] There are also „natural" latent solution tests, when one is studying several unconnected populations of a species and finds the behaviour in question to be invented in at least two populations (e.g., ant dipping in chimpanzees, Schöning et al., 2008; Whiten et al., 2001).

other primates and birds), some of which have focused on the tool-use domain (for a review, see Reindl et al., in press; see also Bandini & Tennie, 2017; Bandini, Neadle, & Tennie, in prep.). These tests provide evidence that several tool-use behaviours in orangutans (Lehner, Burkart, & van Schaik, 2010; Bandini et al., in prep.) as well as algae scooping in chimpanzees (Bandini & Tennie, in press) do indeed represent latent solutions.

The ZLS theory claims that many (maybe all) animal species possess a ZLS, i.e., behavioural traits that can be acquired individually and do not rely on social learning – including humans[3]. However, in contrast to humans, the material cultural behaviours of great apes (and possibly all other non-human animals) are argued to be limited by the respective species' ZLS (*ZLS-only hypothesis*, Reindl et al., in press). This is because these animals seem to be unable to acquire traits (i.e., behaviours, knowledge, skills) that are "novel" to them, i.e., traits that they could not have invented individually but that rely on social learning to be acquired (Tennie et al., 2009; see also Köhler, 1925; Tennie, Call, & Tomasello, 2012). Tennie et al. (2009) acknowledge that this is a strong claim that could be refuted. However, as long as there is no counterevidence, the strong claim will be upheld, also to encourage researchers to carry out experiments on animals' social learning capacities.

Yet, as indicated above, the ZLS-only claim does not rule out the existence of cultures because social learning can affect the frequency of individual reinventions of a given behaviour in a population. Indeed, other authors also suggested that many animal cultures are based on a combination of individual learning and "simple" social learning via second-order conditioning, stimulus or local enhancement (Alem et al., 2016; Heyes, 1993, 2012a; Hill, 2010; Logan, Breen, Taylor, Gray, & Hoppitt, 2015; but see the case of vocal

[3] Note that the size and content of the ZLS can differ between species.

imitation in cetaceans and birds for possible exceptions, Boyd & Richerson, 1996; Rendell & Whitehead, 2001; Slater, 1986).

1.2.1.2 A ZLS in humans?

Whereas great ape cultural behaviours are suggested to fall within the respective species' ZLS (ZLS-only claim, Reindl et al., in press; Fig. 1, left side), it is evident that a large proportion of human (tool) culture (e.g., computers, electric hammers, cars or ships) could not have been invented by any individual without access to previous cultural knowledge – thus representing the product of the ratchet effect (Basalla, 1988; Boesch & Tomasello, 1998; Henrich, 2015; Tennie et al., 2009; Tomasello, 1999b; Petroski, 1997). Human groups differ with regard to the size and complexity of their tool repertoires, but even those with the seemingly simplest repertoires rely on the use of tools that are the result of cumulative cultural evolution (Henrich, 2004; McGrew, 1987). For example, after rising sea levels cut Tasmania from Australia 10,000 years ago, the Tasmanian toolkit lost much of its size and complexity so that by the arrival of the Europeans it had become the simplest known human toolkit. Yet, the Tasmanians were still using fire and manufactured tools consisting of more than one functional part (*technounit*) – aspects of tool use that are not seen in even the most complex chimpanzee tool cultures and that probably constitute cumulative culture (McGrew, 1987)[4].

[4] However, note that these aspects of tool use could also be human latent solutions. Whether they are latent solutions or rather represent cumulative culture would have to be investigated empirically.

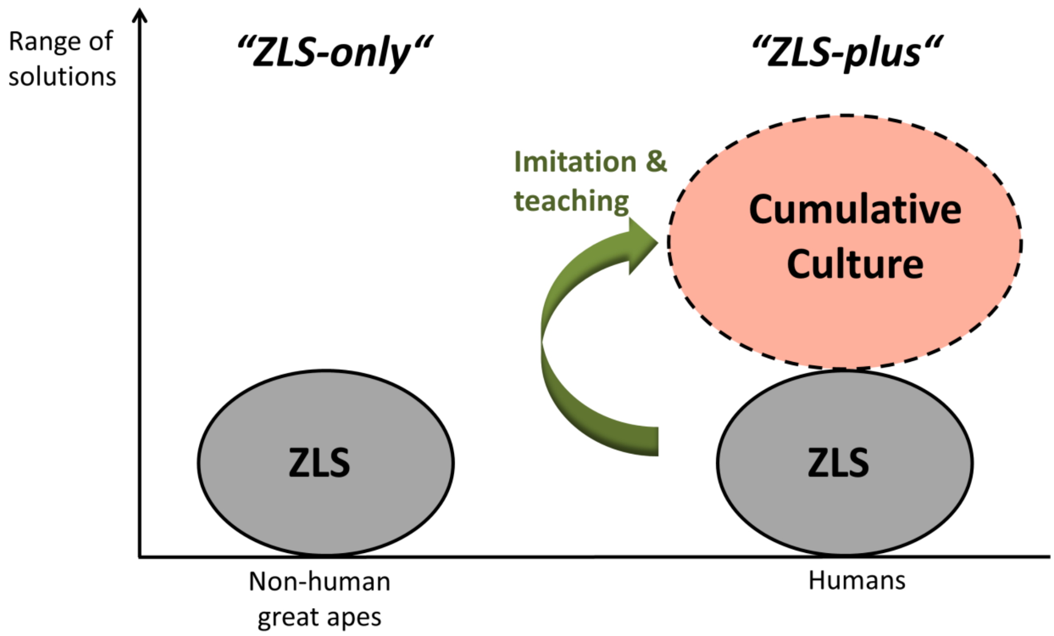


Fig. 1. The zone of latent solutions (ZLS) theory applied to the case of non-human great apes and humans. The ZLS of humans and great apes are of equal size in this figure, but the theory acknowledges the possibility that they might differ in size and content.

The capacity to produce cumulative culture – i.e., to accumulate beneficial changes in cultural traits beyond what individuals can reach on their own (Boyd & Richerson, 1996) – has enabled humans to substantially extend their range of solutions beyond the scope of their ZLS (*ZLS-plus hypothesis*, Reindl et al., in press; Fig. 1, right side). Our ability to produce cumulative culture is argued to rest largely on our high-fidelity social transmission mechanisms, such as *imitation*, i.e., the copying of a demonstrator's goals, actions, and results of these actions (Carpenter & Call, 2002), and imitation-based teaching (Hoppitt et al., 2008; Tennie et al., 2009; Tomasello, 1999a, 2016). These mechanisms allow for faithful transmission of skills and knowledge across generations so that subsequent learners do not have to "reinvent the wheel" but can copy the already existing cultural traits and can invest their cognitive capital into the enhancement of these traits.

On first glance, the ZLS account seems to emphasize differences in the cognitive basis of human and great ape cultures: It claims that whereas most – or even all – cultural behaviours in great apes lie within their ZLS, large parts of human cultural traits need faithful social learning to be acquired. However, the theory also suggests an important similarity across species: the existence of a ZLS (Fig. 1). It assumes that humans, too, possess a range of tool behaviours that do not depend on high-fidelity social learning. This assumption is in contrast with the view of one of the classics of psychology, Lev Vygotsky, who suggested that all human tool use was acquired by imitating others, and that young children's spontaneous tool use was "practically zero" (Luria & Vygotsky 1930, p. 114). However, this view would imply either that human and great ape tool use is qualitatively different with no common evolutionary history ("humans imitate, great apes (re)invent") or that at some point after the split of the hominin and great ape lineages at 14 million years ago humans – but no other great apes – had lost all their spontaneous tool-using abilities. Yet, this seems unlikely given the enormous adaptive value of tool-use behaviours (Laland et al., 2000). In addition, it seems implausible that human cumulative culture (e.g., tool behaviours that need to be transmitted via high-fidelity social transmission mechanisms) has emerged "from zero" (Reindl et al., in press).

Therefore, the ZLS account suggests that in the tool-use domain, humans are not born as blank slates but that there is a phylogenetic baseline of simple, *individually acquirable* tool behaviours (the ZLS) upon which more sophisticated forms of tool use that are acquired through *high-fidelity social learning* are added (cumulative culture, via the collective brain, Fig. 1). We need to identify those tool behaviours that lie within the human ZLS for several reasons: Not only would it establish more evolutionary continuity with regard to tool-use abilities in humans and other great apes, it would also help identify which human tool behaviours actually rely on social learning. In addition, it would allow species comparisons with regard to the content of the ZLS. Lastly, by applying the methods of cognitive cladistics (e.g., identifying which tool behaviours are shared by humans and great apes, Byrne 1995) we can use the gathered data to make suggestions about which tool behaviours the last common ancestor of humans and great apes could have possibly invented without high-fidelity forms of social learning.

1.2.1.3 Exploring the human ZLS

In order to identify tool behaviours within the human ZLS, the same logic applies as the one for non-human animals: testing target-behaviour naïve individuals in latent solution experiments, in which they are provided with all the raw material necessary to invent the behaviour in question. For this, one would need to test humans with as little previous cultural knowledge about the target behaviour as possible in order to rule out – or at least reduce – the possibility that participants show the correct behaviour by drawing on previous socially learned knowledge rather than by inventing it through individual trial-and-error-learning or insight. However, carrying out human latent solutions tests is tricky as humans usually experience rich cultural environments from early on. There have been occasional attempts to deprive humans from any cultural influences, particularly from language input ("The forbidden experiment", Shattuck, 1980), but reports from these tests remain anecdotal; more importantly, these experiments are highly unethical and so should never be repeated. An ethically valid approach is testing very young children with novel tasks in order to minimize the probability that participants can draw on much directly relevant previous cultural knowledge.

Research on children's spontaneous tool-use abilities is sparse, and explicit latent solution tests with children have not been carried out yet prior to this thesis. We know from Piaget's (1952) studies that infants start to use tools at the end of their first year: 9- to 11-month-olds are able to pull cloths, strings, and hooks to obtain out-of-reach toys; however, infants still require spatial contact between the tool and the target to succeed (E. Bates, Carlson-Luden, & Bretherton, 1980). The ability to use tools develops gradually through the second year (Rat-Fischer, O'Regan, & Fagard, 2012) and at two years, children can use tools on spatially separate objects (Brown 1990; Chen, Siegler, & Daehler, 2000; Rat-Fischer et al., 2012). Finally, some 3-year-olds have been found to spontaneously make and use hooks from pipecleaners to retrieve rewards from a tube (Beck, Apperly, Chappell, Guthrie, & Cutting, 2011; Nielsen, Tomaselli, Mushin, & Whiten, 2014; Sheridan, Konopasky, Kirkwood, & Defeyter, 2016)[5]. These data already hint at the existence of a human ZLS.

In order to further investigate the range of spontaneous tool use in children, chapters 2 and 3 present the first explicit latent solution tests on human tool use, with the former focusing on children's ability to invent simple tool behaviours that are based on behaviours observed in wild great apes (assumed to lie within the great apes' ZLS) and the latter focusing on children's ability to use two tools in combination (*associative tool use, ATU*).

[5] Note that the fact that some 3-year-olds made a hook in the Beck et al. (2011) study is interpreted differently by one of the authors: Rather than assuming that hook making is in the human ZLS, Beck (pers. comm.) argues that these children had seen similar acts before. This interpretation cannot be ruled out, of course.

1.2.2 A powerful extension to the human ZLS: Cumulative culture

Humans are able to extend their behavioural repertoire via the production of cumulative culture – culture that no individual could have invented on his/her own. This ability is argued to rest on two key cognitive capacities: high-fidelity social transmission and innovation (Richerson & Boyd, 2005). These are accompanied by other cognitive, motivational, and demographic factors such as prosociality, conformity, normativity, the size and connectedness of a group, and the ability to filter out maladaptive traits (Caldwell & Millen, 2009; Dean, Kendal, Schapiro, Thierry, & Laland, 2012; Dean et al., 2014; Derex, Boyd, 2015; Enquist, Ghirlanda, Jarrick, & Wachtmeister, 2008; Henrich, 2004; Kempe & Mesoudi, 2014; Lewis & Laland, 2012; Muthukrishna & Henrich, 2016; Powell, Shennan, & Thomas, 2009; Tomasello, 1999a, 2009). Another cognitive factor has recently been re-emphasized by the proponents of the *cultural attraction theory* who state that individual reconstruction – i.e., the mental re-building of an observed trait by the individual, during which the trait is influenced by the individual's cognition (e.g., preferences, motivation, memory capacity) and by previous knowledge, resulting in the individual "construct[ing] a variant of her own" (Claidière & Sperber, p. 91) rather than "just" copying the observed action or result of that action – can account for the stability and spread of many cultural traits (Claidière, Scott-Phillips, & Sperber, 2014; Morin, 2016a; Sperber, 1996).

While it still remains to be seen how these factors interact and which of them are strictly necessary for cumulative culture, a strong case has been made for a crucial role of high-fidelity learning through imitation and imitation-based teaching (Richerson & Boyd, 2005; Tomasello, 1999a; but see Morin, 2016a): From an early age, when observing others, humans exhibit a strong focus on action compared to results information, which allows them to copy actions faithfully (Call, Carpenter, & Tomasello, 2005; Horner & Whiten, 2005; Nagell, Olguin, & Tomasello, 1993; Tomasello, 2009; Whiten, Custance, Gomez, Teixidor, & Bard, 1996). This ability is argued to be especially important for the acquisition of 1) complex instrumental skills that involve novel action sequences and/or that contain actions whose causal relevance is not apparent to the learner (e.g., complex tool making) and 2) purely-action based, conventional acts such as gestures, language, rituals, and dances (social-conventional culture; Gergely & Csibra, 2006; Heyes, 2013; Legare & Nielsen, 2015; Moore, 2013; Tennie et al., 2012). In contrast, great apes do not seem to spontaneously copy actions as they either tend to pay more attention to endstates than to actions and/or lack the motivation to copy observed actions (Call et al., 2005; see also Clay & Tennie, 2017; Tennie, Call, & Tomasello, 2006; but note that some enculturated great apes do sometimes imitate after extensive training; Tomasello, 1999a; Whiten, Horner, Litchfield, & Marshall-Pescini, 2004).

Recent evidence suggests that our capacity for imitative learning is not present from birth (Oostenbroek et al., 2016; see also Heyes, 2016a), but that it develops within the first year of life (Ray & Heyes, 2011)[6]. At 1 year, infants flexibly switch between imitation (achieving the same goal as a demonstrator by using the same actions and producing the same results) and *emulation* (achieving the same goal and producing the same results as a demonstrator, but using one's own behavioural means; Nielsen, 2006), and so young children possess a portfolio of social learning mechanisms (Whiten, McGuigan, Marshall-

[6] The extent to which this development relies on cultural learning vs specialized biological adaptations is still debated (Heyes, 2012b; Tomasello, 1999a).

Pescini, & Hopper, 2009). Therefore, from early on, children are equipped with the cognitive and motivational prerequisites for acquiring cumulative culture.

However, only few studies so far have investigated at which age children start to acquire cultural traits that they could not have invented on their own (see chapter 4). Importantly, no study has examined when children become able to acquire such traits in the context of material culture. It might be apparent from everyday experience that children are able to learn new things that are beyond the scope of their individual learning capacities (how else would children be able to learn anything new?). However, these capacities nevertheless need to be studied scientifically. Chapter 4 contributes to the literature two studies investigating whether children can learn from and copy material cultural traits they could not have invented on their own.

The theoretical groundwork for this question was laid out in the 1920s by Lev Vygotsky (1978) who introduced the concept of the zone of proximal development. The next section will therefore describe his theory and its connection to cumulative culture and the ZLS.

1.2.3 The ontogenetic roots of human culture: The zone of proximal development

Given the definitions of cumulative culture – behavioural traits that *no member of a species* could invent individually (Boyd & Richerson, 1996) – and of the ZLS – the range of behaviours that *any able-bodied member of a species* can potentially invent individually from scratch (Tennie et al., 2009) – it becomes apparent that both concepts operate on a species level. Vygotsky, however, was interested in the individual, studying how individual humans extend their current behavioural repertoire by social learning. For this, he introduced two concepts: The *zone of actual development (ZAD)* describes an individual’s current behavioural repertoire, i.e., the behaviours that the individual at a given point in development can carry out independently, without help from others (Fig. 2). Individuals can differ in the size and contents of their ZAD: For example, two 10-year-olds – despite being the same age – might differ in the complexity of arithmetic calculations they can do on their own. The ZAD can grow and gradually incorporate more skills. Those skills that are about to be integrated into the ZAD are labelled as the zone of proximal development (ZPD, Fig. 2). The ZPD encompasses the skills and behaviours that an individual at a given point in development is able to achieve only with the help of more knowledgeable others. The ZPD is limited: There are always skills that are still beyond the ZPD and thus cannot be learned yet, even if help (e.g., teaching) is available (Fig. 2). Again, just like with the ZAD, individuals can differ in the size and content of their ZPD.

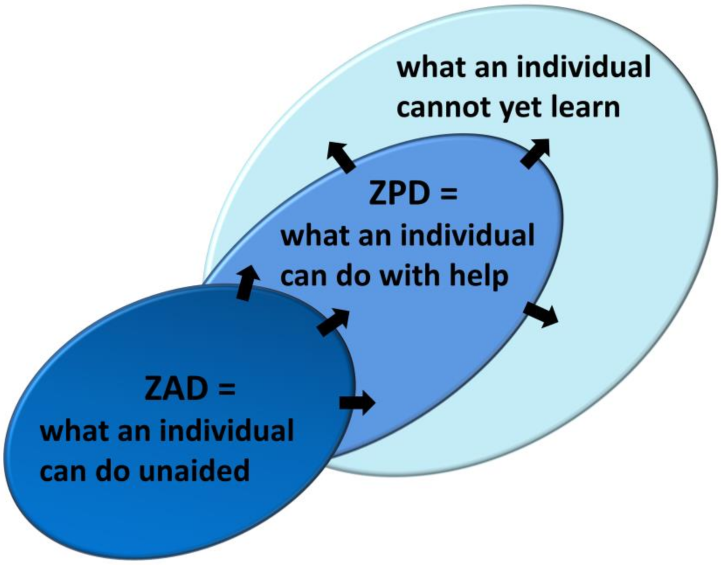


Fig. 2. Vygotsky's (1978) zone of actual development (ZAD) and zone of proximal development (ZPD). Arrows indicate that the respective zones can grow and incorporate new traits.

Vygotsky (1978) argued that human learners can extend their ZAD because of their capacity for imitation and because of demonstrators' capacity for teaching ("scaffolding")[7]. Note the analogous role that social learning plays on the species level: Tennie et al. (2009) assume that the human species can go beyond its ZLS because of its capacity for imitation and imitation-based teaching. Importantly, however, the ZAD and the ZLS are not the same because 1) they operate on different levels (individual vs species) and 2) the ZAD can include behaviours that need to be learned socially, whereas the ZLS by definition does not. Yet, the ZLS and ZAD are complementary: The ZLS represents the "phylogenetically derived "baseline" of the ZAD" (Reindl et al., in press), the "first and

[7] Thus, Vygotsky presented a precursor of the ontogenetic cultural intelligence hypothesis (Herrmann et al., 2007; Tennie & Over, 2012; van Schaik & Burkart, 2011; see also Reindl et al., in press).

non-cultural instalment of the human ZAD, upon which a potential for development through social learning exists (ZPD)” (ibid.).

The ZLS-plus account (Tennie et al., 2009) and Vygotsky’s (1978) ZAD and ZPD concepts can be combined to provide a framework for studying (cumulative) culture from both an individual and a species perspective (Reindl et al., in press). In Fig. 3 I visualize my understanding of how the ZLS and the ZAD and ZPD concepts intertwine in the human case[8]. The starting point are the coloured “zones” in the middle of the figure. The areas below, in the respective colours, contain the definitions of the respective zones. The dark blue field in the middle (ZAD_i) represents the ZAD of individual *i*, i.e., those traits that *i* can potentially invent on his or her own. The ZAD_i at a given point in time t_1 consists of 1) already acquired ZLS behaviours (*ZLSi*;), i.e., behaviours whose acquisition did not require social learning, and 2) behaviours which had previously been within the ZPD_i and which were incorporated into the ZAD_i through social learning (*former* ZPD_i).

[8] Appendix 1 contains the corresponding figure for the case of great apes.

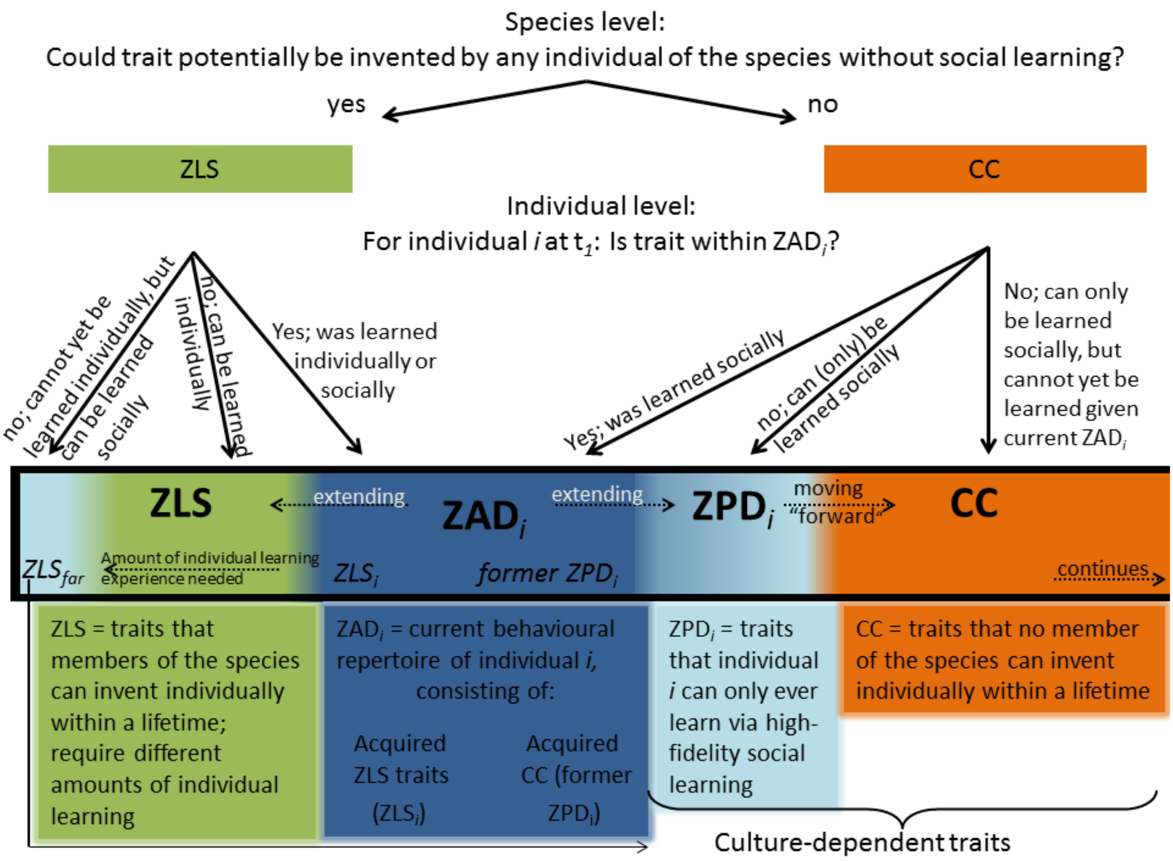


Fig. 3. An integration of Tennie et al.'s (2009) zone of latent solutions (ZLS) theory with Vygotsky's (1978) zone of actual development (ZAD) and zone of proximal development (ZPD), and the concept of cumulative culture (CC).

The ZLS, depicted as the green area, forms the phylogenetic basis for the ZAD_i. Note that ZLS behaviours do not automatically lie within an individual's ZAD. Rather, humans start their life with a relatively small ZAD and need to incorporate new skills through learning processes. Think for example of a 1-year-old infant: As I will show in chapter 2, there is a range of simple tool behaviours which lie within the human ZLS. However, infants' cognitive and motor skills are not sufficiently developed yet for them to invent any of these behaviours on their own. Therefore, the infants' ZAD would not yet contain these behaviours, even though they are within the human ZLS. Those latent solutions that an individual at t_1 cannot yet master are labelled ZLS_{far} (light blue area on the left). ZLS_{far} traits do not need social learning in order to be acquired; however, I assume that some of them could be acquired already at t_1 when social learning is available[9]. Supporting evidence for this thought comes from studies showing that children at a given age who find a given behaviour very difficult or impossible to invent on their own (e.g., infants using a rod to obtain an out-of-reach toy) are able to learn it when receiving a demonstration, while the acquisition in children who did not receive a demonstration occurs only later in development (Chen et al., 2000; Beck et al., 2011; Rat-Fischer et al., 2012; Somogyi, Ara, Gianni, Rat-Fischer, Fattori, O'Regan, & Fagard, 2015).

The ZPD_i – the bright blue area on the right – contains traits that individual *i* at a given point in time can only acquire through social learning (this is in contrast to traits within the ZLS that do not require social learning at any point during the individual's development). The ZPD always represents a subset of cumulative culture (*CC*). CC is depicted as the orange, potentially infinite area at the right and contains traits that no human could invent individually. Cumulative cultural traits that are within the individual's immediate "developmental level" (Vygotsky, 1978, p. 85), lie in the ZPD_i and could be incorporated next into the ZAD_i, provided that an opportunity for high-fidelity social learning is available.

The lines in the bottom part of the figure demonstrate the definition of *culture-dependent traits*: Culture-dependent traits are those traits that can only be acquired through high-fidelity social learning, i.e., traits within the ZPD_i, the ZLS_{far}, and CC. The arrow going from the ZLS_{far} zone symbolises that ZLS_{far} traits are culture-dependent traits at a

[9] Other ZLS behaviours, however, might not be learnable at t_1 even when social learning is available, for example in cases in which still developing motor skills or physical strength are limiting factors.

given point in time for a given individual, but may not remain so with the individual's development (e.g., at t_2 the individual will be able to invent the trait independently).

The top part of Fig. 3 further demonstrates how the ZLS theory and Vygotsky's theories complement each other: For any given trait, one could first ask or hypothesise whether this trait represents a latent solution or cumulative culture for a given species. For example, for humans, knowledge about Bayesian statistics likely represents cumulative culture, as it was generated and refined over several generations. Then, on an individual level, one can ask whether the trait has already been mastered by the individual. For example, Eric might not yet know much about Bayesian statistics, but he might already have acquired some knowledge about statistics in general so that he would be able to follow and learn from a course on Bayesian statistics – Bayesian statistics would be within his ZPD. However, for four-year-old Max, who just learned how to count, Bayesian statistics would be beyond his ZPD because he has to master many other steps in between; for Max, this would place Bayesian statistics into the CC zone of Fig. 1. This framework highlights the need for differentiation between the species and the individual level. For example, studying a trait that (supposedly) represents cumulative culture for a species, one has to bear in mind that the trait could be beyond the current learning capacity for an individual (CC), or that it could lie just within the ZPD_i and could thus be acquired through high-fidelity social learning, or that it could be a part of the individual's current behavioural repertoire (ZAD_i).

What is interesting with regard to the second half of this thesis is the case where individual *i* represents a child rather than an adult. Chapter 4 investigates the origins of cumulative cultural learning in humans by studying children's capacity to learn from and copy culture-dependent traits. For this it is crucial to demonstrate to children a culture-dependent trait that falls within the ZPD of children at that age (rather than lying beyond it). For example, one cannot demonstrate to preschool children the Eiffel tower and expect them to grasp the underlying physics and to build such a tower themselves. That is, it is important to not present children with anything that is likely to lie even beyond the ZPD of many adults. Instead, we need to present children with a trait of which we know that children of that age could not invent it on their own but of which we are confident that children could copy it when provided with demonstrations (e.g., a simpler tower from sticks and plasticine instead of the Eiffel Tower). This implies that the demonstrated culture-dependent trait might have to be so "simple" that it actually is not an example of human cumulative culture, but that it actually is a ZLS_{far} trait, i.e., a trait that children at a given age can only acquire through social learning but that humans would nevertheless be able to invent at some point during their lifetime even without social learning. Determining whether a trait (e.g., a simple stick-and-plasticine tower) represents "true" cumulative culture or a ZLS_{far} trait is difficult because at the point when individuals are able to invent the trait individually (i.e., when they are older), it is practically impossible to rule out that socially acquired cultural knowledge was necessary for the invention. Although classifying a trait as CC or ZLS_{far} is a challenge on the theoretical level, it should be noted that this is extraneous to the study of children's acquisition of culture-dependent traits as long as it has been established that the demonstrated trait cannot be invented individually by children in the studied age range.

The theoretical framework presented here might prove useful in preventing potential confusion and misunderstandings among researchers. For example, with regard to our study on children's cultural learning in which we presented young children with simple towers (chapter 4), we received feedback from several researchers asking in how far these towers represent cumulative culture. Importantly, they do not. What they do represent, however – and what we demonstrate in chapter 4 – is a cultural trait that children can only acquire through social learning, i.e., a culture-dependent trait. (ZLS_{far} in Fig. 3).

1.3 Summary and aims of this thesis

This introduction outlined and interconnected the ZLS theory, cumulative culture, and Vygotsky's concepts of the ZAD and ZPD. Combining these concepts provides a useful, cross-disciplinary theoretical framework for the study of (human) culture, but also points to existing gaps in our knowledge and encourages research efforts. The ZLS theory focuses on the origin of cultures, suggesting that cultures – including human culture – start to emerge from a series of individual reinventions (latent solutions) influenced by social learning rather than from social learning per se. And so the theory predicts the existence of a ZLS for humans. The first two empirical chapters of this thesis are dedicated to testing this hypothesis. They present two explicit latent solution tests for human tool-use abilities, aiming to identify a series of simple tool-use behaviours that young children are able to invent spontaneously and unaided. Specifically, chapter 2 presents a published paper investigating whether 2-year-olds are able to invent a range of tool-use behaviours that are also shown by two of our closest living relatives – chimpanzees and orangutans. This study also sheds some light onto the phylogenetic origins of our tool culture, suggesting that the studied tool behaviours are shared by humans, non-human great apes (hereafter great apes), and probably also by their last common ancestor. Chapter 3 examines whether young children are also able to spontaneously invent more complex types of tool use, namely using two tools in combination (ATU), with most of these tasks again being inspired by behaviours observed in wild or captive animals.

Chapters 4 and 5 add another aspect of studying the developmental basis of human material culture: They move away from investigating children's *spontaneous* tool-use abilities and focus on children's capacity to *socially learn* from, copy, and improve material culture. One hundred years ago, Vygotsky (1978) noted that humans are able to learn novel cultural traits through imitation and teaching. However, explicit tests as to when and how children learn novel material cultural traits that they could not have invented on their own (i.e., culture-dependent traits) have not been carried out yet. Thus, chapter 4 presents two studies that investigate for the first time whether 4- to 6-year-olds are able to copy a culture-dependent material product and which social information children need to do so. Using the same task, chapter 5 examines whether groups of children are already able to produce culture-dependent material products themselves by learning from and innovating upon each other's ideas. Chapter 6 summarizes the studies, relates them to existing research, highlights their limitations, and gives an outlook onto possible future pathways for research.

To address our research questions, we ran various behavioural experiments with young children. Our general approach for statistical analysis of the data was to establish – a priori – a specific version of a *Generalized Linear Model (GLM)* which also contained a random effects structure if repeated measurements had to be accounted for (*Generalized Linear Mixed Model (GLMM)*). The models were used to investigate the potential relationship between the response variable and several predictor variables. Models containing a random effects structure were constructed in a way that they contained not only all random effects that needed to be accounted for, but also all possible random slopes, following the suggestion by D. J. Barr, Levy, Scheepers, and Tily (2013). Where these "full models" did not converge due to overparameterization, we created an a priori, theory-driven model simplifaction process. Descriptive analyses were carried out in SPSS, GLMs and GLMMs were run in R. In the chapters, the models are referred to as either ANOVAs, regressions, or a version of a GL(M)M; the variety of labels is due to our increase in experience over time which led us to go away from more traditional labels such as ANOVA and regression. We left these original labels as the respective papers were also published with this nomenclature. However, it should be noted that all these procedures are closely related as they are built from the same basic linear model.

Ethical approval for all studies was granted by the University of Birmingham, UK, STEM Ethical Review Committee. All studies were carried out in accordance with the approved guidelines.

CHAPTER 2: YOUNG CHILDREN SPONTANEOUSLY INVENT WILD GREAT APES' TOOL-USE BEHAVIOURS

This chapter is a modified version of the paper:

Reindl, E., Beck, S. R., Apperly, I. A., & Tennie, C. (2016). Young children spontaneously invent wild great apes' tool-use behaviours. *Proceedings of the Royal Society B, 283.* doi: 10.1098/rspb.2015.2402.

For this chapter, the main text and the supplementary material of the paper have been rearranged to allow for better readability. Minor modifications have been made throughout the text, but otherwise the text is as published.

I am the primary author of this publication. The original idea for this study was developed in collaboration with my supervisors Claudio Tennie, Sarah Beck, and Ian Apperly. I was primarily responsible for the design of the studies and I carried out all data collection and analysis. My supervisors contributed to authorship by providing feedback and editing versions of this paper leading to its publication.

Link to open access article:

http://rspb.royalsocietypublishing.org/content/283/1825/20152402

2.1 Introduction

The ability to use tools, i.e., to employ “unattached or manipulable attached environmental object[s]” (Shumaker, Walkup, & Beck, 2011, p. 5), is not restricted to humans. Chimpanzees and orangutans – two of our closest living relatives – possess multiple tool-use “traditions”, i.e., tool-use behaviours whose occurrence cannot be explained solely by genetic and environmental factors, but which are also influenced by social learning (Whiten, 2005). Although these traditions bear superficial similarities to human culture (Whiten et al., 2001), the range and complexity of human tools are unique. Exploring the reasons for this uniqueness, researchers have focused on the role of special types of social learning (Boyd & Richerson, 1996; Nagell et al., 1993; Tomasello, 1999a, 2001; Want & Harris, 2002). As Vygotsky (1978) argued, humans’ capacity for imitating and teaching others enables them to acquire behaviours which they cannot (yet) invent on their own. This is because a learner’s capacity for attending to the actions of a demonstrator, a sensitivity to pedagogical cues such as eye gaze, and an ability to faithfully copy these actions allows acquiring a close copy of the behaviour (often even when the purpose and/or goal of the actions are opaque to the learner) and avoids the need to individually reinvent (parts of) the behaviour, as would be the case with other social learning mechanisms such as stimulus enhancement. Thereby, the learner can acquire novel behaviour that was previously outside of her individual reach; she will start at a higher knowledge level than the previous generation rather than having to “reinvent the wheel”. Over historical time, this ability allowed humans to gradually accumulate beneficial design changes in their (tool) cultures (Boyd & Richerson, 1996; Tomasello, 1999a). In contrast, since the evidence for imitation and imitation-based teaching in great apes is weak (Boesch, 1991, 2012; Nagell et al., 1993; Tennie et al., 2012, but see Whiten et al., 2009), tool use in these species is unlikely to be acquired via these mechanisms. The myriad of human tool forms thus represents the current end-result of cumulative cultural evolution. This, however, begs the question as to what types of tool-use behaviours humans can invent without social learning. In other words: What are the roots of our tool cultures – both phylogenetically and ontogenetically?

This chapter explores this "baseline" of human tool-use abilities by asking which tool-use behaviours human children are able to invent on their own, i.e., without cultural resources such as instructions, demonstrations or eavesdropping. In order to determine whether a given tool-use behaviour can be invented individually, culturally naïve individuals can be tested for spontaneous reinventions of the behaviour via latent solution tests (see chapter 1; Tennie & Hedwig, 2009). Whereas for non-human animals this can be done with captive individuals that happen to be naïve to the behaviour in question (Menzel, Fowler, Tennie, & Call, 2013), the case is more challenging for humans, as children learn how to use cultural tools such as spoons from an early age (Connolly & Dalgleish, 1989; McCarty, Clifton, & Collard, 2001). By designing novel tasks, which children are unlikely to have encountered before, researchers can limit human participants' likelihood of drawing analogies from their previous cultural experience when solving the task. Thus, we presented children with novel games and unusual apparatuses to ensure that they would need to solve the tasks via spontaneous individual inventions.

We based our tasks on cultural tool-use behaviours that have been observed in wild chimpanzees (Whiten et al., 2001) and orangutans (van Schaik et al., 2009). This allowed us to identify candidate ecologically valid tool-using behaviours that may be invented spontaneously by children. In addition, such an approach allows for a cross-species comparison of spontaneous tool-use abilities. This is important as one cannot a priori assume that humans possess equal, worse or better basic tool-use skills than great apes – this must be tested explicitly (Herrmann et al., 2007). Flexible tool use, i.e., tool use that is not a stereotyped behavioural adaptation but that requires some amount of individual learning to be acquired (Hunt, Gray, & Taylor, 2013), is thought to require a range of physical cognition capabilities: Tool users (of any given species) need to understand that objects can be used as tools (Hunt et al., 2013), to recognize the tools' functionality (Ruiz & Santos, 2013), and they need causal reasoning skills (Hunt et al., 2013). In addition, more general abilities seem to be required, such as recombining information to solve novel tasks (Call, 2013), inhibition to switch between strategies, the ability to learn from perceptual-motor feedback (Hunt et al., 2013), and an enhanced working memory capacity to process the increased *problem-solution distance* of tool-use tasks (i.e., tool tasks often exhibit a greater distance in time and/or space between the recognition of a problem and its solution; the start and end states of tool-use acts are often separated by several, e.g., detect food – search for suitable tool – modify tool – take tool to target location – use tool – obtain food; Haidle, 2010; Hunt et al., 2013). Finally, a propensity for object manipulation (Call, 2013) is also required. Deaner, van Schaik, and Johnson (2006) found that some non-human primates possess better domain-general cognition abilities than others and thus speculated that these differences in general intelligence might account for differences in these species' tool-use abilities. Following this logic, one might expect that humans – possessing superior domain-general cognition abilities than other primates – might *excel* other primates on tool-use tasks. Alternatively, Ruiz and Santos (2013) argue that humans and great apes possess a *similar* understanding of the physical and functional aspects involved in tool use (but differ in their understanding of the social aspects). Another approach might expect human tool-use abilities to be *impoverished* compared to other species: Modern human intelligence is argued to be based on our species-unique cultural learning abilities (cultural intelligence hypothesis; Boyd & Richerson, 1996; Henrich, 2015; Tomasello, 1999a, 2011). Such a reliance of human intelligence on social transmission could have led to the loss of some individual cognitive skills[10]. Indeed, in many domains our cultural history has resulted in humans being born less well equipped than great apes. For example, humans have lost their fur because they developed clothing (a culturally transmitted tool) and cooking has shrunk our digestive organs, making us dependent on this cultural form of "pre-digesting" (Henrich, 2015; Wrangham, 2009, Zink & Lieberman, 2016). Therefore, our cultural intelligence could have made individual "baseline" physical cognition obsolete, suggesting that in our tool-use tasks human children could be outperformed by great apes.

The physical cognitive skills of children and great apes have been compared in one previous study that presented chimpanzees, orangutans, and 2.5-year-old humans with a test battery of physical and social cognition tasks (Herrmann et al., 2007). While it suggested that children had more advanced social skills than great apes, no differences were found in their physical cognition abilities. However, the tasks used in this study were solely inspired by human behaviour and so they lacked the ecological validity from the perspective of the great apes – which are arguably closer to the state of our common ancestor than modern humans (Wrangham, 2001). Thus, using great ape behaviours as the basis for tasks for humans might represent a phylogenetically more appropriate approach if

[10] Note that the suggestion that humans could have lost some of their individual cognitive skills does not necessarily imply that humans also lost their capacity for innovation. As Muthukrishna and Henrich (2016) argue, innovations often result from collective intelligence, i.e., groups of individuals recombining their knowledge and skills, rather than from special individual cognitive skills.

we aim to make inferences about our last common ancestor's cognitive capacities. Our approach, by creating tasks for humans based on great ape behaviours, now complements Herrmann et al.'s (2007) method, and in combination both represent a more valid approach to the comparative analysis of human physical cognition.

The first aim of the study was to conduct the first explicit latent solution test in humans by investigating whether children between 2 and 3.5 years would be able to spontaneously invent tool-use behaviours required to solve naturally-occurring problems that wild great apes solve. For this we created the *Great Ape Tool Test Battery (GATTeB)*, containing 12 tasks derived from cultural tool behaviours observed in wild chimpanzees and/or orangutans. To be able to conclude that a behaviour lies within the spontaneous capacities of children (and thus likely within the human ZLS), we would need to observe its spontaneous invention in at least two participants (Huffman & Hirata, 2004; see methods).

If we found that at least some of the tasks could be spontaneously solved by the children, i.e., if we found an overlap in the spontaneous tool behaviours of children and great apes, this would suggest – from a *phylogenetic* perspective – that these behaviours were also likely within the spontaneous cognitive reach (the ZLS) of our last common ancestor. With regard to the *ontogenetic* basis of human tool use, our findings would add to the sparse literature on spontaneous tool behaviours in children. The ability to spontaneously use tools (such as sticks or cloths) to obtain out-of-reach objects has been shown to develop between 8 and 24 months of age (Rat-Fisher et al., 2012). However, whether young children are also able to spontaneously use tools for other purposes, such as extracting or perforating objects, is still unknown. Thus, our findings would contribute to the developmental literature insights into whether children would also be able to spontaneously use tools for these purposes that have not been investigated, and which are ecologically highly relevant.

We also aimed to investigate whether tool-use behaviours which appear to be more difficult to invent for wild great apes would also be more challenging for children to invent spontaneously. Cultural tool behaviours in wild great apes differ with regard to their observed frequency: While some behaviours are shared by several communities within a species, e.g., termite-fishing in chimpanzees, others are less frequent, e.g., chimpanzees' use of a stick for algae scooping (Whiten et al., 2001). It has been suggested that the behaviours' observed frequency represents their ease of individual invention (Tennie et al., 2009). Thus, the tool behaviours occurring rarely in wild great apes may pose greater demands on cognitive abilities (e.g., planning, working memory) and/or motor skills (e.g., physical strength, fine motor skills) when compared to more frequent tool behaviours. Would behaviours which appear to be hard to invent for wild great apes also be more difficult to invent for children in our study? To investigate this, we divided the 12 GATTeB tasks into two groups (low-frequency, high-frequency), according to the frequency with which the respective great ape behaviours were observed in the wild.

We studied children from 2 to 3.5 years. We chose the lower end of our age range to be 2 years as Herrmann et al. (2007) have claimed that 2-year-olds represent a meaningful point of comparison with great apes. However, we chose a broader age range than Herrmann et al. for two reasons: First, our tasks represented more challenging tool problems, with many of them posing additional demands on children's planning, fine motor skills or physical strength. Second, by allowing for variation in participants' age we were able to examine whether any difference between low- and high-frequency tasks was stable over developmental time.

2.2 Methods

2.2.1 Creation of the Great Ape Tool Test Battery (GATTeB)

We based our test battery on tool-use behaviours described in the current reviews of potentially cultural behaviours in wild chimpanzees (Boesch, 2012; Langergraber et al., 2001; Whiten et al., 1999, 2001) and orangutans (van Schaik et al., 2003, 2009; Table 1). First, we extracted the behaviours from these reviews and identified 124 different behavioural patterns; behaviours that occurred in both chimpanzees and orangutans were counted only once (Table 2; behaviours that were counted only once are printed in bold). Behavioural variants describing similar tool-use actions and functions were merged: *Marrow pick*, *Eye eat*, and *Brain eat* (described in chimpanzees) were comprised to one variant labelled *Marrow pick. Termite fish*, *Ant-fish*, *Ant-dip*, and *Grub extraction* (chimpanzees) as well as *Tree-hole tool-use (*orangutans) were combined to *Termite fish/Tree-hole tool-use. Nut extract* (chimpanzees) and *Seed extraction* (orangutans) were combined to *Seed extraction/Nut extract.* Finally, *Branch drag/Drag branch* (chimpanzees) and *Branch dragging display on ground* (orangutans) were combined to *Branch drag*. Second, we deselected all non-tool-using behaviours, all instances of ATU (to be investigated separately in chapter 3), as well as all universal behaviours.

Table 1. *Chimpanzee and orangutan communities listed in the respective reviews.*

Species	**Population**	**Reference**
Chimpanzee	Bossou (Guinea) Budongo (Uganda) Gombe (Tanzania) Kibale Kanyawara and Kibale Ngogo (Uganda) Mahale B, Mahale, K, and Mahale M (Tanzania) Taï North, and Taï South (Ivory Coast)	Boesch (2012), Langergraber et al. (2011), Whiten et al. (1999, 2001)
	Fongoli (Senegal) Goualougo (Congo) Loango (Gabon)	Boesch (2012)
	Lopé (Gabon) Mt. Assirik (Senegal)	Whiten et al. (2001)
Orangutans	Gunung Palung (Borneo) Tanjung Puting (Borneo) Kutai (Borneo) Lower Kinabatangan (Borneo) Leuser Ketambe (Sumatra) Leuser Suaq Balimbing (Sumatra)	van Schaik et al. (2003, 2009)
	Sabangau (Borneo) Tuanang (Borneo)	van Schaik et al. (2009)

Table 2. *Excluded behavioural variants in chimpanzees and orangutans and reasons for their exclusion.*

Reason for exclusion	Behavioural pattern: Chimpanzees	Behavioural pattern: Orangutans
	No tool use	*No tool use*
No tool use	1 Buttress-beat[a]	19 Water play[c]
	2 Branch-clasp[a]	20 Coercive hand-holding[c]
	3 Muzzle rub[b]	21 Throat scrape[c]
	4 Skull pound[b]	22 Twig biting[c,d]
	5 Driver ant-hand[b]	23 Symmetric scratch[c,d]
	6 Termite mound-pound[b]	24 Nest smack[c]
	7 Herbal pith[b]	25 Raspberry[c,d]
	8 Ground-day-nest[b]	26 Snag riding[c,d]
	9 Ground-night-nest[a]	27 Drink from bottom of pitcher plant[c]
	10 Day cushion[b]	28 Slow loris eating[c,d]
	11 Day nest[b]	29 Dead twig sucking[c,d]
	12 Rain dance[a]	30 Nest destruction[d]
	13 Knuckle-knock[a]	31 Bouquet feeding[c,d]
	14 Hand-clasp[a]	32 Long-call vibrato[c]
	15 Index-hit[a]	33 Kiss-squeak with hands[d]
	16 Leaf-groom[a]	34 Copulation on female‘s nest[c]
	17 Food-pound onto wood[a]	35 Using Asplenium fern to rest or sleep in[c]
	18 Food-pound onto other[a]	36 Females rubbing their genitals together[c,d]
		37 Biting through vine to swing across gap[c]
		38 Biting through vine to release tree to sway to adjacent tree[c,d]
		39 Washing face and arms with water from tree hole[c]
		40 Male and female use the same nest to spend the entire night[c]
		41 Play nests[d]
		42 Nest as social refuge[c]
		43 Snag crashing[d]
		44 Sneaky nest approach[c,d]
		45 Hide under nest[c,d]
		46 Artistic pillows[d]
		47 Bridge nest[d]
		48 Carry leafy branch to different tree to build nest[d]
		49 Leaf bundle[c,d]

Reason for exclusion	Behavioural pattern: Chimpanzees	Behavioural pattern: Orangutans
	Tool use	*Tool use*
Universal behaviours	50 Branch-shake[a] 51 Play-start[a] **52 Drag branch/ Branch drag[a,b]** 53 Leaf sponge[a,b] 54 Investigatory probe/Inspect stick[a,b]	
Associative tool use	55 Sponge push-pull[a] 56 Anvil prop[a,b] 57 Open and probe[a,b]	
Hygiene behaviours	**58 Leaf-napkin**[a,b] 59 Comb[a] 60 Nasal probe[a]	**58 Leaf napkin[c,d]** 61 Use leaf to clean body surface[c,d] 62 Moss cleaning[c] 63 Nail cleaning[c] 64 Tooth cleaning[c] 65 Tooth pick[c] 66 Chewing leaves into pulp then smearing foam over body[c]
Handling ectoparasites	67 Leaf-squash[a] 68 Leaf-inspect[a]	
Wound care	**69 Leaf-dab[a,b]** 70 Leaf wadge[b] 71 Wound inspect[b]	**69 Poultice use[c]**
Aggressive behaviour	72 Club/stick club[a,b] 73 Aimed throw/ Missile throw[a,b] 74 Flail twig[b]	
Sexual behaviour		75 Autoerotic tool[c,d]
Communicative context	76 Leaf-clip, mouth[a] 77 Leaf-clip, fingers[a] 78 Leaf-strip[a] 79 Branch din[a] 80 Branch-slap[a] 81 Shrub-bend[a]	**52 Branch dragging display on ground[c,d]** 83 Kiss-squeak with leaves[c,d] 84 Leaf-wipe in kiss-squeak context[d,c] 85 Hiding behind detached branch from predators/humans[c,d]

Reason for exclusion	Behavioural pattern: Chimpanzees	Orangutans
	82 Stem pull-through[a]	
Provoking reactions from other species	**86 Fly-whisk[a,b]** 87 Bee probe[a,b] 88 Expel/stir[a]	**86 Branch as swatter[c,d]**
Spontaneous behaviour	89 Self-tickle[a,b]	90 Scratch stick[c,d]
Possible early cultural influence in humans	91 Seat stick[a] 92 Stepping-stick[a]	93 Branch cushion[c] 94 Leaf gloves/cushions[c,d] 95 Cover head with leafy branch/leaves Against stinging bees[c]
Leaves/ twigs not allowed in nurseries	96 Container[a] 97 Leaf-wipe[a,b] 98 Leaf-brush[a] 99 Leaf mop[a] 100 Seat vegetation[a] 101 Brush-stick[a] 102 Resin-pound[a]	103 Leaf scoop[c,d] 104 Branch scoop[c.d] 105 Sponging[c,d]
Potentially harmful to children		106 Use gloves to get into ants' nest or to handle spiny fruits[c]
Nest building behaviours		107 Bunk nests[c,d]
No clear behavioural description	108 Dig[a]	
Not transferable to laboratory setting	**109 Branch hook/hook stick[a,b]** 110 Leaf rain cover[b] 111 Pestle-pound[q,b]	**109 Branch hook[c]** 112 Sun cover[c,d]

Note. Behaviours in bold print are the same in both species and thus counted only once.

[a]Behaviour listed in Whiten et al. (1999, 2001) and Langergraber et al. (2011). [b]Behaviour listed in Boesch (2012). [c]Behaviour listed in van Schaik et al. (2009). [d]Behaviour listed in van Schaik et al. (2001).

We assessed the remaining behaviours with respect to transferability to problem-solving tasks for children, deselecting those that failed to transfer: We deselected nine behaviours related to great ape hygiene; two aiming at handling ectoparasites; three related to wound care; three shown in an aggressive context; and one sexual behaviour. We excluded 10 behaviours used in a communicative context and three behaviours because they aimed at provoking reactions from other animals, which was difficult to emulate in the lab: *Fly-whisk/Branch as swatter*, *Bee probe*, and *Expel/Stir*. Two behaviours were excluded as they were dependent on rare and spontaneous incidents that were difficult to provoke in the laboratory (*Self-tickle* and *Scratch stick*). Four behaviours were deselected because we agreed that children were likely to possess cultural knowledge about them: *Seat stick* and *Branch cushion* (see cushions), *Stepping-stick* (see shoes), and *Cover head with leafy branch to protect against stinging bees* (see hats). Ten behaviours were excluded as we were not allowed to take leaves or twigs into nurseries. *Use gloves to get into ants' nest or to handle spiny fruits* was deselected as it was regarded as potentially harmful to children. One behaviour was excluded because it was related to nest building behaviour *(Bunk nests)*. *Dig* was excluded because we did not find a clear behavioural description; its differentiation from *Lever open* seemed to be unclear. Four behaviours were dropped because they were not practicable: *Branch hook/Hook stick* would have required providing an out-of-reach stick-like object that would possibly have to be attached at the ceiling – this task would have been more suitable for a laboratory environment instead of nurseries; *Leaf rain cover* and *Sun cover* are actions elicited by circumstances which cannot be controlled in an experimental setting (sun, rain); and *Pestle-pound* would have required children to destroy something with much physical force, which would have met the space and safety requirements in the nurseries.

The remaining 12 tasks formed the basis for the GATTeB and were divided into two groups according to their observed frequency in the wild (Table 3). We decided to combine those behaviours which were classified by the reviews as *customary* (i.e., occurring in (almost) all members of at least one age-sex class), *habitual* (observed repeatedly in more than one individual, but not customary) or *present* (clearly identified, but not customary/habitual; Whiten et al., 1999) in three or more distinct populations of a species into a *high-frequency* group. Behaviours described as 1) being a *rarity* (i.e., behaviour too rare to spread socially, van Schaik et al., 2009) or 2) being present, customary or habitual for no more than two distinct great ape populations were assigned to a *low-frequency* group. In order to count as "distinct", populations had to be separate, i.e., not be regarded as connected as it is the case, e.g., in the chimpanzee groups of Mahale (Mahale B, K, and M) or Taï (Taï North and South). To give an example, chimpanzees in Bossou as well as in both Taï groups have been observed to use stones or wooden clubs as hammers to crack open nuts. Since the two Taï groups experience exchange of group members, which influences the likelihood of individual reinvention of nutcracking, the groups are likely not independent of each other and we therefore considered them as a single population. Consequently, Nuthammer was counted for two independent groups only (Bossou and Taï) and thereby falls into the *low-frequency* group.

Table 3. *Classification of the tool-use behaviours based on their frequency in the wild.*

Behaviour	**Frequency in the wild**	**Frequency group**
Insect-pound (IN)[a]	rare (present in Bossou)	Low-frequency
Perforate (PER)[a,b]	Habitual/customary in one population (Goulaougo)	
Nuthammer (NUT)[a,b]	Habitual/customary in two connected populations (Taï North and South) and present in another (Bossou)	
Algae scoop (AE)[a,b]	Customary in one population (Bossou)	
Ground puncture (GR)[b]	Customary in one population (Goualougo), at least present in another (Fongoli)	
Seed extraction/Nut extract (SEED)[b,c,d]	Orangutans: Customary in one population (Suaq Balimbing); chimpanzees: present in two connected populations (Taï North and South)	
Marrow pick (MA)[a,b]	Customary in two connected populations (Taï North and South), at least present in another two (Gombe, Goualougo)	High-frequency
Fluid-dip (FD)[a,b]	Customary in three populations (Taï, Lopé, Gombe), habitual in four (Assirik, Mahale K and M, Kibale)	
Ant-dip-wipe (ADW)[a]	Customary in one population (Gombe), habitual in one population (Assirik), present in one population (Bossou)	
Termite-fish leaf-midrib (TFLF)[a]	Customary in one population (Mahale K), habitual in one population (Assirik), present in one population (Bossou)	
Lever open/stick as chisel (LEV)[a-d]	Chimpanzees: customary in two populations (Gombe, Lopé), habitual in one (Taï); orangutans: rare (Ketambe and Tanjung Puting)	High-frequency
Termite-fish/Tree-hole tool-use (TF)[a-d]	Chimpanzees: customary in two populations (Gombe, Mahale K), habitual in one (Assirik); orangutans: customary in one population (Suaq Balimbing)	

[a]Behaviour listed in Whiten et al. (1999, 2001) and Langergraber et al. (2011). [b]Behaviour listed in Boesch, C. (2012). [c]Behaviour listed in van Schaik et al. (2009). [d]Behaviour listed in van Schaik et al. (2001).

Nine of the 12 behaviours were derived from behaviours only shown by chimpanzees. Three tasks (Seed extraction/nut extract, Termite-fish/tree-hole tool-use, Lever open/stick as chisel) were based on behaviours that occur in a comparable fashion in both chimpanzees and orangutans. They were assigned to the frequency groups in the following manner: As Seed extraction and Nut extract occurred with low frequency in both orangutans and chimpanzees, we assigned Seed extraction/Nut extract to the low-frequency group. For Lever open and Termite fish we found that these behaviours were highly frequent in chimpanzees, but the respective versions in orangutans were of only low frequency. We decided to list both variants as highly frequent, as they occurred with high frequency in at least one species.

Although the selected great ape tool behaviours are all exhibited within a foraging context, we did not use food as a reward for human participants due to ethical issues. Instead, each task was designed as a game in which children could win a sticker. Stickers represent a highly valuable and desirable good for most Western children throughout the preschool age – and are thus motivating for children. Table 4 presents an overview of the 12 selected behaviours as well as their adaptation as tasks for children. Fig. 4 depicts photos of the tasks.

Table 4. *Selected great ape tool-use behaviours and description of the GATTeB tasks.*

Behaviour (Frequency)	Description of great ape behaviour	Description of task	Allocated testing time
Insect-pound (low)	Use stick to pound bottom of hole to break and retrieve insects	Use stick to retrieve Play Doh balls from tube by prodding them	2 min
Perforate (low)	Use stick to make probing holes in termite nests	Use stick to perforate barrier in box to retrieve sticker	2 min
Nuthammer (low)	Use piece of wood/stone to crack nuts	Use clay hammer to crack plastic nut to obtain sticker	2 min
Algae scoop (low)	Use twig to scoop for algae on water surface	Use stick to scoop for strip of plastic in polystyrene beads to obtain sticker	2 min
Ground puncture (low)	Use stout stick to puncture underground insect nest	Use stout stick to puncture layer of plasticine in box to retrieve sticker	3 min
Seed extraction/nut extract (low)	Use twig to extract seeds from nut/fruit	Use stick to extract pom poms from box	2 min
Marrow pick (high)	Use small stick to retrieve marrow of long bones	Use stick to retrieve sponge attached to sticker from tube	1 min
Fluid-dip (high)	Use sticks to fish for honey or water	Use stick to dip for paint in tube	1 min
Ant-dip-wipe (high)	Use stick to collect ants, then wipe off and eat	Use wet stick to collect polystyrene beads, then wipe off into box	3 min
Termite-fish leaf-midrib (high)	Use leaf midrib to retrieve termites from nest	Subtract paper “leaf” from stick and use stick with Velcro at ends to fish for scourer pieces in box	2 min

Behaviour (Frequency)	Description of great ape behaviour	Description of task	Allocated testing time
Lever open/stick as chisel (high)	Use stick as lever to enlarge insect nest entrance in log or ground	Use stick as lever to enlarge hole in plasticine lid of a mug to retrieve ball with sticker attached to it	1 min
Termite-fish/tree-hole tool-use (high)	Use stick to extract insects from nest	Use stick with Velcro at ends to fish for scourer pieces in box	1 min

Low-frequency tasks	High-frequency tasks
 a) Insect-pound	 g) Marrow pick
 b) Perforate	 h) Fluid-dip
 c) Nuthammer	 i) Ant-dip-wipe
 d) Algae scoop	 j) Termite-fish leaf-midrib
 e) Ground puncture	 k) Lever open/stick as chisel
 f) Seed extraction/nut extract	 l) Termite-fish/tree-hole tool-use

Fig. 4. Apparatuses used for the GATTeB tasks. Panels 1a) to 1f) present the low-frequency tasks, panels 1g) to 1l) the high-frequency tasks.

2.2.2 Participants

Fifty children (24 boys) between 26 and 41 months (*M* = 33.04 months, *SD* = 3.69 months; 12 26- to 30-month-olds, 30 31- to 36-month-olds, eight 37- to 41-month-olds) were tested individually in nurseries (*n* = 33), a Science Museum (*n* = 9) and our Infant and Child Laboratory (*n* = 3) in Birmingham, UK, as well as in a nursery in a small town in southern Germany (*n* = 5). The ethnic background of the sample was mostly Caucasian (68%), 26% of the children were Black and 6% Asian.

Children in nurseries were recruited via information letters sent to parents after initial contact with the nurseries had been established. Those tested at the Science Museum were recruited via an advertisement on the museum website or were approached directly in the museum. Children tested in the laboratory were recruited via an existing database. The testing situation was comparable across the testing sites: The experiment took place in a separate room or quiet corner of a room in the nurseries or the science museum. Children were tested individually by the same female experimenter (E), and received the same warm-up procedure.Written informed consent was obtained by children's guardians prior to the study. Each participant was administered four tasks (two high- and two low-frequency tasks, randomly chosen and put in one of the following two orders: high-low-high-low or low-high-low-high). Each task was administered to between 15 and 17 children.

An additional 22 children were tested but had to be excluded from the analysis because they were too shy (*n* = 2), cried (*n* = 1) or did not match the required age range (*n* = 19). The children who did not match the age range were all older (up to 52 months), but were tested as in some nurseries information on children's age was given to us only after data collection was completed.

2.2.3 Material

2.2.3.1 Warm-up task

Children were presented with an A4 picture of a meadow with horses and a smaller one of a farmer. The game was to help the farmer build a fence by breaking a rectangular stick of Balsa wood (l = 15 cm, b = 0.5 cm) into shorter pieces.

2.2.3.2 Low-frequency tasks

Insect-pound (IN)

A vertical, green plastic tube (l = 10 cm, diameter = 2.2 cm) was glued to a piece of cardboard (l = 19 cm, b = 10 cm). The tool consisted of a terracotta-coloured stick made of air-hardening modelling clay (l = 24 cm, diameter ~ 1 cm), which had three tiny wooden spikes (l = 2 mm) at each of its ends. These were the blunt ends of skewers which had been inserted into the still wet clay. The spikes enabled the user to retrieve three balls of Play Doh (diameter = 1 cm) from the tube by pounding and prodding them.

Perforate (PER)

The apparatus consisted of two parts: A round, transparent plastic box (diameter = 8 cm, h = 4 cm) containing the reward (a sticker glued to a die) and a red cardboard box (14 x 11 x 6 cm) glued on top of the plastic box. The only entrance to the plastic box was via a hole (diameter = 4.5 cm) at the top of the red box. There was a small slit (l = 8.5 cm, h = 1 cm) at the bottom of the red box in which a round piece of flower arrangement foam (diameter = 8 cm, h ~ 0.8 cm) was inserted. This piece blocked the entrance to the plastic container. A green, wooden stick (l = 19 cm, diameter = 0.6 cm) served as the tool.

Nuthammer (NUT)

The anvil consisted of a cardboard (37 x 27 cm) with a soft foam surface on the left, a hard papier-mâché surface on the right (both: l = 18 cm, b = 22 cm) and a container (l = 29 cm, b = 4 cm) for the nut at the rear part of the board. Both surfaces showed a depression (diameter ~ 7 cm, h = 1.5 cm) in the middle where the nut could be placed. A sticker was placed inside a brown plastic sphere (diameter = 3.5 cm) which consisted of four equal parts made by a 3D printer and put together with a water soluble glue. A lump of clay (l = 10.5 cm, b = 5 cm) was used as the hammer.

Algae scoop (AE)

A red cardboard box (30 x 23 x 11 cm) with a transparent lid and two openings in an inversed-T shape (horizontal part: l = 19 cm, h = 1 cm; vertical part: l = 3.5 cm, h = 3 cm) at the right and left side was used in this task. The box was filled with white polystyrene balls (diameter 3-5 mm) which were used as practical alternative to water. The reward consisted of a sticker attached to a black piece of light plastic foil (l = 20 cm, b = 3 cm). A wooden yellow and blue stick (l = 28 cm, diameter = 0.6 cm) served as the tool.

Ground puncture (GR)

The apparatus consisted of two plastic boxes (20 x 17.5 x 12 cm) which were glued together on top of each other and covered by cardboard and colourful wrapping paper to appear like one single box (25 x 18 x 24 cm). The bottom of the upper box contained a hole (l = 10.5 cm, b = 9 cm) which was covered by a layer of blue plasticine (h = 1-1.5

cm). Two sticks inserted from the side of the box as well as a cardboard frame on top of it fixed the plasticine to the bottom of this box. The box on the bottom had round windows (diameter = 8 cm) on both long sides so that the reward inside was visible. The reward consisted of a sticker glued to a yellow piece of cardboard (l = 4 cm, b = 4 cm). A blue wooden stick (l = 51.4 cm, diameter = 2.2 cm) with a pointed end was used as the tool.

<u>Seed extraction/Nut extract (SEED)</u>

A blue papier mâché box (18 x 7 x 2.5 cm) with a narrow opening at the top (l = 11 cm, b = 1 cm) was used for this task. Six pom poms (diameter = 1.5 cm) in different colours were used as the target objects which had to be removed from the box by levering. A red and blue wooden stick (l = 19 cm, diameter = 0.6 cm) served as the tool.

2.2.3.3 High-frequency tasks

<u>Marrow pick (MA)</u>

Children were presented with a transparent test tube (l = 15 cm, diameter = 3.7 cm). The reward consisted of a sticker attached to a rolled up piece of sponge (l = 5 cm, b = 4.5 cm) inserted at a depth of 5.5 cm from the top (indicated by a black line). A clear Perspex stick (l = 22.7 cm, diameter = 0.5 cm) was used as a tool.

<u>Fluid-dip (FD)</u>

The apparatus consisted of a test tube ((l = 15 cm, diameter = 3.7 cm) inserted into a yellow cardboard box (18.5 x 9 x 7 cm) and a red bottle lid (diameter = 3 cm). A small amount of yellow children’s paint was used as the target object. A wooden stick (l = 19 cm, diameter = 0.6 cm) served as the tool.

Ant-dip-wipe (ADW)

A transparent oval container (13 x 10.5 x 6.5 cm) with a hole (diameter = 2 cm) in the lid was glued on top of a water bottle (h = 27 cm). The bottle contained 1 liter of water and a piece of sponge with a small hole served as a lid. The oval container was filled with white polystyrene balls (diameter 3-5 mm). The apparatus was placed in a blue cardboard box (25 x 22 x 39 cm) with only 2 cm of its upper part visible. A blunt glass stick (l = 49.5 cm, diameter = 0.7 cm) was used as a tool. When removing the stick from the apparatus, the polystyrene balls would stick to its wet surface. Another object in this task was a transparent plastic container (26 x 23.5 x 12.5 cm), which children used to put the polystyrene balls in.

Termite-fish leaf-midrib (TF-LF)

The apparatus consisted of a yellow cardboard box (21.5 x 9 x 12 cm) with a hole (diameter = 3.5 cm) in it. Three pieces of green sponge scourer (l = 1.5 cm, b = 1.5 cm) with a star attached to each of them were used as target objects. The tool consisted of a wooden stick (l = 28 cm) with Velcro glued to both ends and a sturdy paper "leaf" (l = 7.5 cm, b = 7.5 cm) at 7.5 cm from one side.

Lever open/stick as chisel (LEV)

A metal mug (h = 11 cm, diameter = 8 cm) covered in colourful wrapping paper was covered by a layer of blue plasticine (h = 1.5 – 2 cm) with a hole (diameter = 0.5 – 0.7 cm) in the middle. A clay ball (diameter = 1.5 cm) with a sticker attached to it was used as the target object and placed into the mug. A yellow plastic stick (l = 13 cm) with a slightly pointed tip served as the tool.

Termite-fish/Tree-hole tool-use (TF)

The apparatus consisted of a colourful sloping cardboard box (21 x 17 x 17 cm) with a hole (diameter = 3.5 cm) in it. Three pieces of green sponge scourer (l = 1.5 cm, b = 1.5 cm) with a star attached to each of them were used as target objects. The tool was built of a wooden stick (l = 28 cm) with Velcro attached to both ends.

2.2.4 Procedure

Participants were sitting at a table or on the floor, facing E or sitting perpendicular to her. A warm-up game in which children had to break wooden sticks familiarised children with the fact that they were allowed to modify and destroy material during the experimental session. This was important as several tasks required participants to break or perforate material and sometimes also involved applying physical force (e.g., Ground puncture). All GATTeB tasks were designed to be solved by using a tool. Based on observations in a pilot study, children were given 1 min (Termite-fish/tree-hole tool-use, Lever open/stick as chisel, Fluid dip, and Marrow pick, as children found the solutions quickly), 2 min (Insect-pound, Nuthammer, Termite-fish leaf-midrib, Perforate, Algae scoop, and Seed extraction/nut extract) or 3 min (Ground puncture (as children had to apply much physical force) and Ant-dip-wipe (as children tended to only slowly approach the apparatus)) to solve the tasks. E gave general encouragement, but never suggested using the tool.

The German and English version of the instructions can be found in Appendix 2. The instructions were phrased in a fashion as general as possible, not suggesting the use of the tool as the solution to the task (e.g., in Perforate, Algae scoop, Ground puncture, and Marrow pick E said "Try to get the sticker out of the box!"). The exception was Ant-dip-wipe where children were told to remove the stick from the apparatus, as in this task we were not interested in children' spontaneous use of the tool per se, but rather on whether they spontaneously used a certain efficient strategy to remove polystyrene balls from the stick, namely holding the tool in one hand while wiping the balls off with one or more strikes of the other hand (as it is the technique in the target behaviour that differs between great ape populations, Schöning et al., 2008).

The rewards following the tool behaviours differ between our human participants and the wild great apes: Whereas children's motivation to solve the GATTeB tasks was to gain stickers, great apes engage in the respective tool behaviours in order to obtain food. However, as Tomasello and Call (2008) pointed out, what is even more important than constructing identical task contexts is establishing functionally equivalent situations and stimuli. And with this regard, our tasks are comparable to the wild great ape tool tasks: Stickers are highly motivating for most young Western children, as food is highly motivating for most wild great apes.

In six GATTeB tasks the sticker was directly involved in the task, i.e., it was placed inside the apparatus and had to be retrieved with the tool. In the other six games, children were told that they could win a sticker if they solved the game (e.g. in Fluid dip, if they were able to obtain some paint from within the apparatus). Children were rewarded with a sticker in each task regardless of success.

2.2.5 Coding

Participants' behaviour was live-coded. We documented whether children picked the tool up or picked it up and used it in any way (*tool pick up/use*), whether they used the tool in a way that could potentially lead to success (*correct tool use*), and whether they succeeded on the task by using the tool in the correct way (*correct success*). The rare cases in which children succeeded in a non-intended way, i.e., without a tool, were scored as incorrect success and were excluded from the analysis. Data from eight children (16% of the sample; the only children for whom video material was available) was coded by a second rater blind to the hypotheses of the study. Cohen's k for tool pick up/use was perfect (k = 1.000), and excellent for correct tool use (k = .874) and correct success (k = .913).

In order to be able to conclude whether a given tool behaviour was within children's spontaneous capacities, we required to observe the spontaneous invention of the behaviour in at least two participants. Positive evidence from a single child was regarded as insufficient because the observed action could have been produced by chance. While it would still be possible that two independent children could have produced the correct behaviour by chance, we think this is highly unlikely. There is always a (small) likelihood that a random behaviour produced by a participant matches the target behaviour by chance. In our case, this chance is likely to be very small because generating the correct solution in any of the GATTeB tasks requires a behavioural sequence consisting of several steps (e.g., in Insect-pound: pick up stick – insert stick in tube – prod Play doh with stick – extract stick). Whereas single elements may indeed be produced or supported by chance, the likelihood that the *sum* of these elements is generated by chance becomes even smaller (for a similar argument see Köhler, 1925). We still acknowledge that a complete behavioural sequence may be invented by chance by one participant per condition. However, the chance likelihoods become even smaller when two or more participants show the correct behaviour because then these small likelihoods have to be multiplied with each other, resulting in likelihoods that are exceedingly small. Therefore, setting a criterion of at least two participants showing a given behaviour allows us to exclude chance as a feasible explanation for double occurrences of a given behaviour (but, again: not for single occurrences).

We started our data analysis with a descriptive analysis of children's rate of tool pickup/use, correct tool use, and correct success. Then we conducted GLMMs in R (R Core Team, 2013) to investigate whether children's performance on these three variables differed between low- and high-frequency tasks.

2.3 Results

Fifty children completed a set of four tasks each, resulting in 200 trials of which 193 were valid. One trial had to be excluded because of an intervention of nursery staff, two trials due to experimenter error. Four trials were excluded after being scored as incorrect success: In NUT, one child succeeded by pounding the ball directly on the ground. Another child managed to open the nut by tearing it apart with his fingers. In AE, one subject was able to insert his hand through one of the openings of the apparatus and to extract the target object. Similarly, in PER, one child managed to put her hand in the apparatus and to break the barrier with her fingers.

Fig. 5 and 6 depict children's rates of tool pickups/uses, correct tool uses, and correct successes for individual tasks and grouped by low- and high-frequency tasks. Out of the 193 times the tasks were conducted with children, the respective tools were picked up in 80.3% of the cases, demonstrating that children were motivated to interact with the tools.

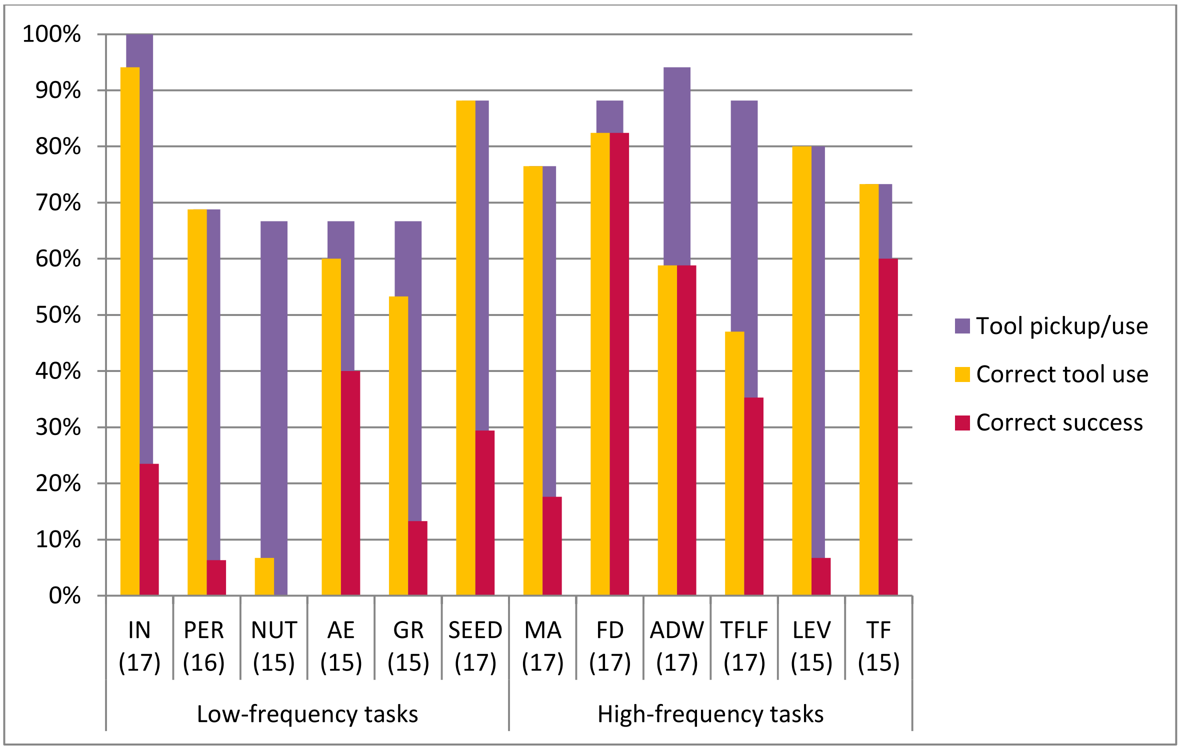


Fig. 5. Rates for tool pickup/use, correct tool use, and correct success for individual tasks. IN = Insect pound, PER = Perforate, NUT = Nuthammer, AE = Algae Scoop, GR = Ground puncture, SEED = Seed extraction, MA = Marrow pick, FD = Fluid dip, ADW = Ant-dip-wipe, TFLF = Termite-fish leaf-midrib, LEV = Lever open, TF = Termite fish. Numbers in brackets are sample sizes.

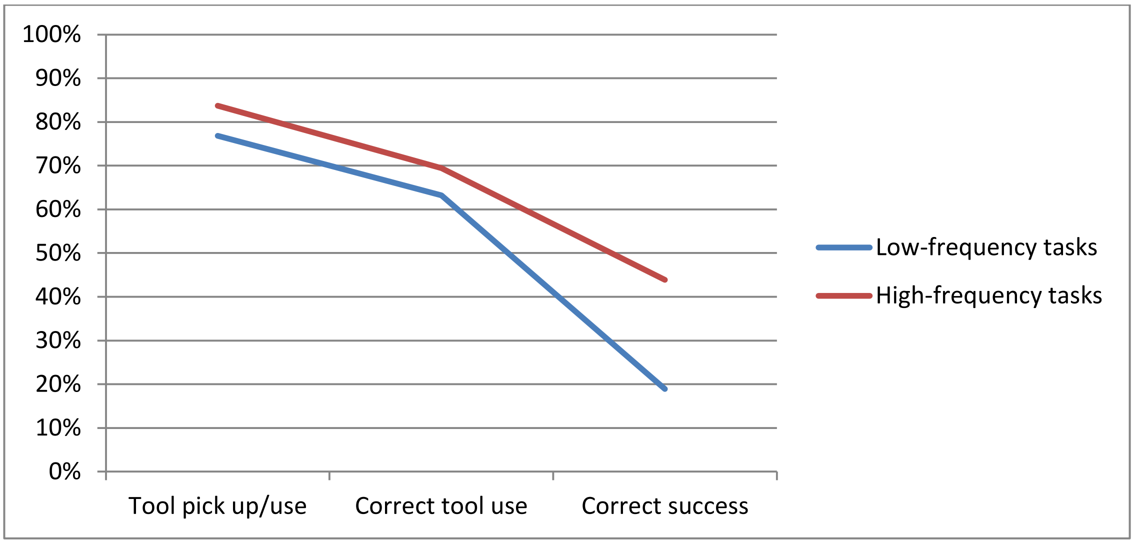


Fig. 6. Rates for tool pickup/use, correct tool use, and correct success for the low- and high-frequency groups.

Correct tool use was observed in 11 tasks (and more than eight times in each of them), indicating that the majority of the great ape tool solutions could be invented individually by human children. “Nuthammer” (i.e., hammering of a plastic nut with a clay hammer) was the only task in which only one child used the tool correctly. Since this child did not succeed in breaking the nut, and since no second child used the tool correctly, we cannot rule out the possibility that this instance of correct tool use was due to chance.

In terms of correct success, we found that in 31.6% of the trials children solved the given task correctly. We also found a large numeric difference in correct success between low- and high-frequency tasks: Whereas children solved 19% of the low-frequency trials, the success rate for high-frequency trials was 44%. In addition, we found that for both low- and high-frequency tasks the majority of successful children (78% in the low-frequency tasks, 70% in the high-frequency tasks) were able to solve the tasks immediately, i.e., they chose to use the tool as their first attempt to solve the task (i.e., they did not use other strategies such as using their hands before they chose to use the tool) and once they picked up the tool, they solved the task "in one go", i.e., without a change of strategies or setting aside the tool in between. Therefore, children's success cannot solely be accounted for by individual trial-and-error learning.

We investigated whether low- and high-frequency tasks differed with regard to their rates of tool pickup/use, correct tool use, and correct success by conducting GLMMs with a binomial error structure and a logit link function in R version 3.0.2 using the glmer function of the R package lme4 (D. Bates, Maechler, Bolker, & Walker, 2013). The aim was to find three models (one for each dependent variable (DV)) which would explain the data best given the predictors age, sex, and frequency. We started by specifying full models including the maximal number of fixed and random effects: Sex, age, and frequency were entered as fixed effects (we also entered an interaction term between age and frequency; as it was not significant, we only report the models with the main effects). The random effects structure was: A random intercept for subjects (to account for the fact that data points were not independent) and a random slope for frequency on subjects, with subjects being nested within the variable *nursery.* The random slope allowed children to respond differently to high- vs. low-frequency tasks. The random intercept and the random slope were allowed to covary.

We used a backward elimination procedure using the "drop1" function to derive the most parsimonious models with the best model fit. We systematically dropped terms that did not significantly contribute to the model fit, i.e., whose removal did not lead to a significantly worse model fit, a change in the significance level of the predictors or a change in the Odds Ratios (OR) of the predictors greater than 10%. We made sure that the models for the three DVs always consisted of the same random effects structure at each step in the elimination process. That is, if we had to remove a random effect in the equation for one of the outcome variables – e.g., because of non-convergence of the model – we did the same for the other two DVs.

For each model we computed p-values, OR, and Confidence Intervals (CI) for the individual predictors. Maximum Likelihood tests were used to derive p-values of the predictors by comparing models including the respective factor with those not including them (using the R function “anova” with the argument “test” set to “Chisq”). Model stability of the three final models was determined by a comparison of the estimates of a model based on the complete data set with those derived from models where the levels of the random factor subject were excluded one at a time. To investigate possible problems concerning multicollinearity, we calculated a standard linear model of the final model which excluded the random effects and determined the Variance Inflation Factor (VIF). We looked for potentially influential cases by using the “DFBETAS” function of the R package influence.ME.

2.3.1 GLMM: Tool Pickup/Use

Since tool pickup/use was not meaningful for ADW (children were told to pick up the tool), we excluded this task for the according GLMMs, resulting in 176 valid trials for these analyses. The initial full model estimated the parameter for the variable frequency with only low fidelity, indicated by a large OR (169) and a wide corresponding CI between 5 and 5451. Therefore, a simplification of the random effects structure was necessary: Excluding the correlation between the random factors, dropping nursery or doing both did not result in a more stable model. Only the exclusion of the random slope for frequency did so. The resulting full model was not able to explain the data better than a null model consisting of only the random effects structure ($\chi^2(3) = 2.363$, $p = .501$).

The “drop1” function suggested that no predictor contributed significantly to the model fit. Since the removal of sex suggested the greatest reduction in the AIC and we did not find any gender effects in the other models, we decided to drop this variable. This did not lead to changes in the significance levels of the predictors and changes in the OR were smaller than 1%.

Although the “drop1” function suggested that age did not significantly contribute to the model fit either, it remained in the equation to allow comparison with the model for correct success, for which the effect of age was significant. Finally, nursery was removed since the variance it accounted for was small. This did not result in any changes in the significance levels or OR of the predictors. This final model still did not explain the data better than a null model consisting only of the random effects structure ($\chi^2(1) = 2.362$, $p = .124$).

We found the model to be fairly stable with regard to frequency and age. There were no problems with multicollinearity in the model (VIFs for both frequency and trial were 1.000). We found nine influential observations. An exclusion of these data points and a recalculation of the model did not lead to a change of the significance levels for frequency. The change in the OR for age was smaller than 1%, the OR for frequency increased by 13%.

2.3.2 GLMM: Correct Tool use

Since we had to drop the random slope for frequency in the model for tool pickup/use, we did the same for the model for correct tool use in order to ensure comparability. The simplified full model was not able to predict the data better than a null model consisting only of the random effects structure ($\chi^2(3) = 2.432$, $p = .488$).

The "drop1" function suggested that no predictor contributed significantly to the model fit. Since the removal of sex suggested the greatest reduction in the AIC and we did not find any gender effects in the other models, we decided to drop this variable. There were no changes in the significance level of the predictors and the changes in the OR were smaller than 1%. Finally, nursery was removed since the variance it accounted for was extremely small. Compared to a null model only consisting of the same random effects structure, the final model did not explain the data significantly better ($\chi^2(2) = 4.142$, p =.126).

We found this model to be very stable with regard to frequency and age. There were no problems with multicollinearity in the model: The values were 1.000 for both frequency and age. We found 14 potentially influential observations. An exclusion of these data points and a recalculation of the model did not lead to changes in the significance levels of the predictors. The change in the OR for age was smaller than 1%, the OR for frequency changed by about 14%.

2.3.3 GLMM: Correct success

Since we had to drop the random slope for frequency in the model for tool pickup/use, we did the same for this model to ensure comparability across the DVs. The simplified full model predicted the data better than a null model consisting only of the random effects structure ($\chi^2(3) = 32.217$, $p < .001$).

The "drop1" function suggested that sex did not contribute significantly to the model fit, so it was dropped from the model. This did not lead to any changes in the significance level or OR of the predictors. Again, we dropped nursery since the variance it accounted for was extremely small. There were no changes in the significance levels or OR of the predictors. The final model explained the data significantly better than a null model just comprising the random effects structure ($\chi^2(2) = 33.377$, $p < .001$). The model revealed a significant positive effect of age ($p < .001$): With each month increase in age, children were 1.3 (95% CI [1.1; 1.4]) times more likely to succeed. On top of this age effect, frequency was a significant predictor for correct success ($p < .001$): Compared to low-frequency tasks, tasks in the high-frequency group were 4.4 (95% CI [2.1, 9.1]) times more likely to be solved (Table 5). We did not find an interaction between age and frequency, i.e., although older children were more successful than younger ones across all tasks, they were still experiencing the low-frequency tasks as more difficult than the high-frequency tasks. The frequency effect was thus stable over the age range. Whether this suggests that the frequency and age effects reflect distinctive or common underlying factors remains open to debate and a focus for future studies.

We found this model to be very stable with regard to age and frequency. The VIFs for age and frequency were 1.000, indicating no problems with multicollinearity in the model. We looked for potentially influential cases by using the DFBETAS function of the R package influence.ME and found 16 observations above the cut-off value. An exclusion of these data points and a recalculation of the model did not result in changes of the significance levels of the predictors. The changes in the OR of the predictors were smaller than 6%.

Table 5. *Final generalized linear mixed model for variable correct success.*

Term	X^2	df	*p*	OR	$CI\text{-}OR_{lower}$	$CI\text{-}OR_{upper}$
Intercept	-	-	-	3.677e-05	4.962e-07	0.003
Age	16.646	1	< .001	1.289	1.139	1.459
Frequency	16.521	1	< .001	4.398	2.124	9.104

Note. Number of observations: 193. OR = Odds ratio; CI = Confidence interval

These results show that great ape low-frequency tasks were more difficult to solve for children and that great ape high-frequency tasks were easier to solve. We found that overall there was a match in frequency categories between apes and children. However, this match was not perfect (Figure 4). For example, some tasks in the low-frequency group had rather high success rates (e.g., Algae scoop), whereas some tasks in the high-frequency group seemed rather difficult for the children (e.g., Lever open). Future studies might benefit from this fact as it might be worthwhile to investigate in detail why some of the tasks did not match this pattern.

2.3.4 Additional analyses

In order to investigate whether children's performance was affected by the fact that tasks were allocated different amount of times for their completion (1, 2 or 3 min), we reran the models including a fixed effect for time. We found that time did not make any significant contribution in the model for tool pickup/use ($\chi^2(1) = 0.555$, $p = .456$). In the model for correct tool use, time had a significant negative effect on children's behaviour ($\chi^2(1) = 6.338$, $p = .012$). With every minute increase in the time allocated to the tasks, children were 0.4 (95% CI [0.2; 0.8]) times less likely to spontaneously use the tools in the correct way. This means that in tasks for which we chose to allocate more time, children were less likely to show correct tool use. This reflects the very reason why we decided to allow children a longer time span for some of the tasks: In a pilot study, where we initially administered only one minute across all tasks, some tasks appeared to be harder for children as correct tool use was observed less often. Thus, we decided to extend the time for these tasks. The results thus reflect that there was a correlation between task difficulty and solution time. For correct success, time did not make any significant contribution in the model ($\chi2(1) = 0.560$, $p = .454$).

In order to rule out the possibility that the difference in performance between the low- and high-frequency tasks could be explained by evident differences in the task design or difficulty, we conducted a small post-hoc study in which we presented 12 adults with the GATTeB and asked them to classify the tasks into two groups containing six tasks each. We scored the number of "correctly" grouped tasks (minimum: 3; maximum: 6) and were thus able to investigate whether subjects would be able to reproduce the classification into low- and high-frequency tasks. Having categorized the tasks, participants were asked to name the criteria they used to split the tasks into two groups.

Three participants grouped three tasks correctly, seven subjects had four tasks correct, and two people classified five tasks correctly. The average number of correctly classified tasks was 3.92 (*SD* = 0.67). We used a Chi-square goodness-of-fit test to compare the distribution of subjects' responses with a distribution resulting from chance classification (expected values drawn from a simulated sample with n = 100,000) and found that subjects were significantly better than expected by chance (χ^2 (2) = 20.35, p < .001). This is not surprising given the fact that the tasks possess some well-perceptible task-inherent similarities which can be used as points of reference for categorization. This is also reflected by the criteria subjects used: Most of the criteria were related to whether the task involved breaking or destruction of objects (e.g., NUT, LEV, GR), obtaining the target object via extraction (e.g., AE, TF, MA) or whether the target object was visible (e.g., MA, GR, AE).

Most critically, however, no participant created a classification equal to the low- and high-frequency split. Furthermore, only one participant mentioned difficulty as a criterion and was therewith able to classify four low-frequency tasks correctly (NUT, GR, AE, IN). However, she claimed that the group containing these four tasks was the easier one, thus not matching the results we found with the children. Thus, our overall conclusion is that although some tasks share some apparent features, these features are not able to distinguish between the low- and high-frequency tasks and thus cannot account for our finding that low- and high-frequency tasks differ in their success rates.

2.4 Discussion

Our study found that the majority (11 out of 12) of the investigated wild great ape tool-use behaviours are individually reinventable by human children and that there is a close relationship between the difficulty level of these behaviours for children and individual discovery rates for great apes. Unlike a previous study (Herrmann et al., 2007) whose tasks were likely biased towards the human case, we validated our tasks ecologically by basing them on great ape tool behaviours as described in the wild. Thus, our study presents phylogenetically more appropriate tasks for the study of the physical cognition of our last common ancestor.

Children showed spontaneous tool use in the majority of the tasks, suggesting that nearly all of the studied behaviours lie within the human ZLS, i.e., within the realm of what humans can invent without observing the solution or having it demonstrated. The large overlap between the behaviours that can be invented spontaneously by great apes and human children suggests that young children's physical cognition skills are at least on the same level as those of great apes. These findings do not rule out the possibility that there might be physical cognition tasks in which young children outperform great apes. However, in combination with the study by Herrmann et al. (2007) – who presented great apes with tasks based on human behaviours and who found no difference in the performance of great apes and 2-year-old children – the results of our study suggest – contra recent claims (Vaesen, 2012) – that ontogenetically, humans do not seem to differ greatly from great apes with regard to their set of physical cognition abilities. From a phylogenetic perspective, humans' basic tool-use abilities do not appear to have become degraded by our species' long reliance on social learning and teaching. However, to eventually answer the question whether the physical cognition abilities of great apes (including humans) are comparable or whether humans possess enhanced physical cognition skills, future studies will need to present humans and great apes with tasks completely novel for both (e.g., tasks based on tool behaviours observed in other, non-primate species and which are not already known to be exhibited by great apes or children).

We found that children were more likely to solve tasks based on great ape behaviours which occur with high frequency in the wild compared to low-frequency tasks, with this effect not changing with age. It seems that tool tasks in the low-frequency group possess features which make successful tool use more difficult for both humans and great apes, i.e., which make them more challenging for the evolved cognition of these species. A possible reason for the enhanced difficulty of the low-frequency tasks might be that whereas high-frequency behaviours mainly require the tool user to perceive and select the correct affordances, low-frequency behaviours may possess additional cognitive or non-cognitive demands. For example, some of the high-frequency behaviours may only require the insertion and retrieval of a stick into a hole (e.g., Termite-fish/Tree-hole tool-use, Fluid dip). In contrast, low-frequency tasks might pose additional demands, e.g., on planning (e.g., Perforate, consisting of two steps: First breaking the barrier with the stick and then turning the box upside down to retrieve the sticker; similarly, chimpanzees need to first break the entrance to the termite mound with a stick and then use a different stick to retrieve the insects), fine-motor skills (e.g., in Seed extraction/nut extract, the target objects have to be retrieved dexterously), physical strength (e.g., in Ground puncture or Nuthammer) or working memory (e.g., in Nuthammer, tool users need to attend to several objects simultaneously). Identifying the specific reasons for the difficulty of low-frequency behaviours will have to be the target of future studies.

Whereas low- and high-frequency tasks differed with regard to children's success rates, we found no effect of frequency on correct tool use. In both low- and high-frequency tasks children were equally likely to show the correct tool behaviour, and did so in more than two thirds of the trials. This finding underlines young children's proneness to use tools in meaningful ways to try to solve even novel problems. However, whether children's disposition to use tools is also followed by task success seems to depend on task type: In high-frequency tasks, both children's tool use and success rates were relatively high. In contrast, in low-frequency tasks, even though children were equally likely to use the tools correctly, tool use was less likely to result in success. This finding highlights that correct tool use does not necessarily imply task success. Other cognitive and/or non-cognitive demands have to be met so that correct tool use can be "translated" into success. This "translation process" seems to be more demanding for the low- compared to the high-frequency tasks (see above for a speculation about possible underlying reasons).

We also found that older children were more likely to solve the tasks than younger children. This suggests a development between 2 and 3.5 years of age of capacities allowing children to more successfully meet the demands of the studied tool tasks. Future work will need to identify these capacities; potential candidates may be improvements in fine-motor skills, visual attention, working memory, physical strength, and planning. We did not find an interaction between age and frequency. That is, even though the older children in our sample might have possessed better planning and fine-motor skills than younger children, this did not suffice to help the older children overcome the demands of the low-frequency tasks. Thus, we conclude that the frequency effect is stable across the studied age range.

Participants only received one trial per task, rendering our approach to studying whether young children would be able to spontaneously invent the necessary tool-use behaviours a rather conservative one. Also, we do not know whether children who produced a correct behaviour in a task would also be able to reproduce it on following trials or whether the behaviour in the first trial occurred only by chance and without insight. Implementing more than one trial would allow children more time and opportunities to learn the correct behaviour individually. In the current study, this would have been especially interesting with regard to Nuthammer, in which only one child produced the correct behaviour. Would this child have been able to use the tool correctly on the following trial, and maybe even have been successful? Future studies administering several trials per GATTeB task are needed to address questions like these.

It might be argued that our tasks were only based on wild tool cultures of two of the four currently living genera of great apes. However, wild gorillas and bonobos exhibit only very low levels of tool use and thus failed to provide the wild input for our tasks. Nonetheless, these genera readily use tools in captivity – i.e., when a need arises to do so (Mulcahy & Call, 2006; Mulcahy, Call, & Dunbar, 2005). Thus, while they did not contribute to our validated list of tasks, they are no exception from the line of widespread tool use across the great apes.

Our findings support the notion that the last common ancestor of humans and great apes – living ~14 million years ago – was already capable of the tool behaviours studied here (and that they also found the low-frequency tasks more difficult to invent). These behaviours thus represent a part of the phylogenetic basic state of human tool use (the ZLS); and they would not have required sophisticated social transmission mechanisms such as imitation and imitation-based teaching.

2.5 Summary and link to chapter 3

This chapter shed more light onto the ontogenetic and phylogenetic roots of human tool culture, identifying a range of ecologically relevant tool behaviours that children –

tested on tasks validated by great ape tool behaviours – can invent on their own. This study is the first explicit latent solution test on human tool use and identified 11 tool behaviours lying in the ZLS of both modern humans and possibly their last common ancestor with the great apes.

The focus of the current study was on “simple” tool behaviours, and so our selection process of tool behaviours from the repertoire of wild great apes left aside an important category of tool behaviours: the use of two or more tools in combination (ATU). To address this gap, the next chapter presents a study conducted in a similar fashion as the current one, but with a focus on children’s ability to spontaneously use *two* tools in combination. Again, we based our tasks – where possible – on spontaneous behaviours observed in wild or captive animals (great apes and birds).

CHAPTER 3: YOUNG CHILDREN SPONTANEOUSLY ENGAGE IN ASSOCIATIVE TOOL USE

"At the present time it is impossible to decide whether the processes which […] appear to us the easiest, have originally evolved most easily and, therefore, earliest. We can only judge what is originally easy, and originally difficult, by means of experimental tests with anthropoids and […] children" (Köhler, 1925, p.65)

3.1 Introduction

3.1.1 General introduction

Chapter 2 investigated whether young children were able to spontaneously invent a range of "simple" great ape tool behaviours, i.e., tool behaviours that involved the use of a single tool. Behaviours comprising the use of two or more tools in combination – a type of tool use labelled associative tool use (ATU; Shumaker et al., 2011) – were excluded from that study. For example, we excluded a behaviour called "Open and probe" in which chimpanzees use a stout stick to open up a termite mound and then use a thinner twig to fish for termites; we also excluded "Sponge push-pull" in which chimpanzees use a stick and a leaf in combination to extract water from a tree-hole. Spontaneous ATU in the wild has been reported for chimpanzees and (Whiten et al., 1999, 2001) as well as for gorillas and capuchin monkeys (Shumaker et al., 2011) for obtaining enclosed or out-of-reach food. ATU also takes up a large part of human tool use. For instance, many humans use knife and fork or chopsticks to eat food, a series of different tools to build and assemble houses and furniture, and toothpaste and toothbrush in combination to look after their mouth hygiene. Crucially, ATU has arguably played an essential role in the evolution of human material culture, especially our ability to use tools to make other tools (*secondary tool use*, Shumaker et al., 2011; Table 6). Given the significance of ATU for human technological culture, I deemed it vital to extend the study of children's spontaneous tool-use capacities to the domain of ATU, as done in this chapter.

Once thought to be a unique characteristic of the human lineage, tool use is now known to represent a behavioural adaptation present in a variety of taxa including mammals, birds, fish, and even insects (Shumaker et al., 2011). This is perhaps unsurprising given the enormous adaptive value of tool use: For example, tools are used to extend one's reach (e.g., using a stick as a hook to haul in a branch with fruits), to amplify forces (e.g., using a stout stick as a lever to enlarge a hole in the ground), to augment display behaviour (e.g., brandishing a stick or branch to intimidate an intruder) or to enhance comfort (e.g., using a stick to scratch oneself; Laland et al., 2000; Shumaker et al., 2011). It has been suggested that tool-using animals can be separated into two groups based on the cognitive mechanisms involved in their tool use (Call, 2013; Hunt et al., 2013): Whereas fish and invertebrates use tools in a mostly stereotyped fashion and in rather specific contexts, so that their tool use can be described as behavioural "specializations", tool use in birds and mammals appears more flexible, i.e., it is influenced by social and/or asocial learning, and is regarded as an expression of general intelligence rather than as a specialization. Since flexible tool use is argued to pose demands on working memory, causal reasoning, and practical skills, it is viewed as a cognitively demanding form of tool use (Hunt et al., 2013, see also chapter 2).

Among these demanding tool-use forms, ATU is arguably the cognitively most challenging type. This is firstly due to the greater number of tools involved (Haidle, 2010;

addition, only humans seem to engage in secondary tool use (Shumaker et al., 2011)[12]. Therefore, researchers have regarded the ability to engage in certain types of ATU as important cognitive characteristic setting humans apart from other animals (Kitahara-Frisch, 1993; Lombard & Haidle, 2012; Vaesen, 2012). ATU has shaped human cognition and culture due to its role in the coevolution of technological advances (complex tool composites, secondary tool use) and cognitive capacities such as working memory or planning abilities (de Beaune, 2004; Read, 2008).

Given this crucial role of ATU for our evolution, it is surprising that research on ATU in humans and other animals is sparse (see below). Yet, studies on ATU capacities can provide important insights into the phylogenetic and ontogenetic origins of ATU as well as into the underlying cognitive mechanisms. Specifically, research on children's spontaneous ATU abilities can identify whether some forms of using two tools in combination represent human latent solutions, i.e., can be acquired through individual rather than social learning (see previous chapters and Tennie et al., 2009). Therefore, the current chapter presents a study whose basic principle follows the logic of the study presented in chapter 2: We first identified and selected ATU behaviours observed in (wild and captive) non-human animals and adapted them to novel problem-solving tasks for children with which we investigated whether children were able to spontaneously engage

---

use has been observed in chimpanzees (e.g., using a stone to prop a stone anvil for nutcracking) and is thus not unique to humans (Shumaker et al., 2011).

[12] Although famous bonobo Kanzi and his sister have been shown to possess the basic capacity for secondary tool use, this was only after extensive periods of teaching and training (Toth, Schick, Savage-Rumbaugh, Sevcik, & Rumbaugh, 1993). Moreover, even after years of practice, Kanzi did not overcome certain cognitive (and anatomical) restrictions to produce tools similar to the earliest hominin stone tools (Toth & Schick, 2009).

Hunt et al., 2013; Read, 2008; Shumaker et al., 2011): A greater number of tools increases the number of relations and interactions between elements in a task (tools are not only acting on target objects, but also on other tools; see *relational complexity account*, Halford, Wilson, & Phillips, 1998). The higher the number of interacting elements that need to be represented in parallel by an individual in order to carry out a task the greater is the *processing complexity* of a task. Tasks with higher processing complexity impose higher cognitive load on working memory and thus demand more complex cognitive processes. A second reason why ATU is likely cognitively rather challenging is because it is usually characterized by an increased problem-solution distance, i.e., the start and end state of an act involving ATU are separated by several steps (Haidle, 2010; Hunt et al., 2013). That is, ATU often consists of a chain of action sequences that increase the temporal and/or spatial distance between the initial problem and its solution (see the *cognigrams* in Haidle, 2010, that impressively visualize the increase of the problem-solution distance when going from "simple" tool use to ATU). A high problem-solution distance poses demands on working memory and other executive functions such as planning, inhibition, sequencing, and decision making (Haidle, 2010).

Whereas ATU is not unique to humans, its ubiquity and level of cognitive complexity in humans seems to be unprecedented in the animal kingdom. Compared to the human case, ATU in non-humans seems to comprise only a small part of their tool behaviours (Shumaker et al., 2011). With regard to complexity, humans are the only species using complex forms of *tool composites* consisting of two separate, but interdependent tools, such as the bow and arrow (Lombard & Haidle, 2012; Table 6)[11]. In

[11] Note the exception of metatool use – a somewhat simpler subtype of tool composites in which one tool (metatool) is used simultaneously with a second tool to increase the efficiency of the second tool. Metatool

in ATU. Before presenting the study, I will describe the different types of ATU currently distinguished in the literature and present previous research on spontaneous ATU in non-human animals and human children[13].

3.1.2 Associative tool use in non-human animals

Depending on the specific way in which two or more tools are used in combination, one can distinguish between different types of ATU. Here, we will use the definitions introduced by Shumaker et al. (2011), but it should be noted that different definitions and nomenclatures are used by different authors (compare for example Shumaker et al., 2011; Taylor, Hunt, Holzhaider, & Gray, 2007; Wimpenny, Weir, Clayton, Rutz, & Kacelnik, 2009). Table 6 presents the different ATU types and their definitions as described in Shumaker et al. (2011). The current chapter only focuses on *tool set*, *metatool use*, and *sequential tool use* (definitions below) and excludes secondary tool use. This is because secondary tool use is argued to be unique to humans and thus might be less suitable for studying the origins of ATU in a comparative manner.

It should also be noted that in contrast to chapter 2, where only behaviours observed in wild great apes formed the basis for our child tool-use tasks, the current study also includes tasks based on behaviours observed in captive animals (sequential tool use). This is because sequential tool use has only rarely been observed in wild animals (so far only in capuchin monkeys, Mannu & Ottoni, 2009) – perhaps due to a lack of necessity. Therefore, in order to find suitable tasks for or current study, we also investigated reports

[13] Reports on ATU in animals that received training will not be included as we are primarily interested in spontaneous, unaided abilities.

on sequential tool use in captive animals (great apes, New Caledonian crows, see below) in order to extend the range of behaviours we could investigate in children. Basing our tasks on laboratory tasks that have been used with other species rather than basing them on behaviours observed in the wild (as done in chapter 2) does not undermine the validity of our study. This is because our primary research question was whether children, *in principle*, are able to invent on their own how to use two tools in combination to solve a novel problem – regardless of whether this problem is completely novel or whether it is part of the problems other species face naturally in the wild. In addition, even though we did not base our sequential tool use tasks on wild animal behaviours, the fact that we did base them on problems that other animals have presented with in the laboratory means that we can still make meaningful comparisons across species with regard to their basic, unaided tool-use skills.

Even though the sequential tool use tasks in the literature might appear somewhat artificial, the study of sequential tool use is as important as the study of other forms of ATU. After all, sequential tool use is a real phenomenon in humans (and capuchin monkeys), so it does have an ecologically valid basis. Also, if the above-mentioned claim is true that ATU is just one form of flexible tool use and so is a mere expression of a species' general intelligence, then one could assume that species that are capable of some type of ATU in the wild (e.g., chimpanzees using tool sets) might equally be capable of showing sequential tool use, even though it has not been observed (yet) in the wild.

Table 6. *Associative tool use (ATU) types and their definitions according to Shumaker, Walkup, and Beck (2011).*

Associative tool use type	Definition
Tool set	Two or more tools used sequentially, usually each in a different mode, to achieve a single outcome
Tool composite	Two or more tools used simultaneously, usually each in a different mode, to achieve a single outcome, where the first tool is not used to manufacture the second.
Metatool use	One tool used simultaneously with a second tool to increase the efficiency of the second tool, where the first tool (metatool) acts directly on the second. The second tool could function as a tool on its own; the metatool makes it a better tool. Every metatool is a tool composite.
Sequential tool use	A tool used to acquire another tool.
Secondary tool use	A tool used to manufacture (structurally modify) another tool.

3.1.2.1 Tool set

A tool set describes the case in which “two or more tools are used sequentially, usually each in a different mode, to achieve a single outcome” (Shumaker et al., 2011, p. 19). A real-world, human example would be using a can opener to open a tin can followed by using a spoon to retrieve the can’s contents (two tools, two functions, one outcome (retrieving food)). In non-human animals, tool sets have been reported for wild chimpanzees and capuchin monkeys: Chimpanzees in several communities across Africa use sets of two or more sticks to get access to beehives, ant nests or termite mounds ("Open and probe"; Hashimoto, Isaji, Koops, & Furuichi, 2015; Shumaker et al., 2011). For example, chimpanzees use stout branches as pounding tools to open up a bee hive and to widen the holes, and then use finer twigs to dip for honey. Capuchins use stones to pound on beehives or next to cavities, and subsequently use sticks to probe for honey or small animals (Mannu & Ottoni, 2009). Captive capuchins were shown to spontaneously use tool sets of stones and sticks to extract the contents of nuts or food containers (Westergaard, Lundquist, Kuhn, & Suomi, 1997).

3.1.2.2 Metatool use

Metatool use is defined as using a tool (metatool) simultaneously with a second tool to increase the efficiency of the second tool (the second tool could also function on its own; Shumaker et al., 2011). An everyday example would be using a cleaning cloth with a cleaning mop handle (see "Swiffer" system). Some wild chimpanzees use wedges to stabilise stone anvils for nut-cracking ("Anvil prop"; Carvalho, Cunha, Sousa, & Matsuzawa, 2008; Matsuzawa, 1991a, 1994). One chimpanzee has been observed to use a stick to push a leaf further down into a tree-hole in order to retrieve water ("Sponge push-pull"; Matsuzawa, 1991b, as cited in Sugiyama, 1997). A similar spontaneous behaviour has been observed in a captive orangutan (Shumaker et al., 2011). Captive capuchins spontaneously used hammer stones (metatool) to pound chisel stones to break open lids or bones (Westergaard, Green, Menuhin-Hauser, & Suomi, 1996; Westergaard & Suomi, 1994).

3.1.2.3 Sequential tool use

Sequential tool use describes "using a tool to acquire another tool" (Shumaker et al., 2011, p. 19). For example, we use a ruler to reach a pen that has fallen behind a desk. In the wild, sequential tool use has been observed in capuchin monkeys who use small stones to loosen larger stones from a rock and then use these large stones to hammer and pulverize pebbles in order to lick the powder (Mannu & Ottoni, 2009). Sequential tool use has been argued to pose significant cognitive load on an individual: It has to understand that a tool can be used on another tool; it needs inhibitory capacity to resist the urge to use the first tool to act directly on the reward instead of getting the second tool; and it needs to be able to hierarchically organize behaviour in order to carry out the task (Bird & Emery, 2009). Studies have shown that New Caledonian crows can use short sticks to retrieve longer sticks suitable for raking in out-of-reach food (Taylor et al., 2010; Wimpenny et al., 2009); depending on the spatial layout of the task, some crows succeeded in their first trial (Taylor et al., 2007). All four great ape species (Köhler, 1925; Martin-Ordas, Schumacher, & Call, 2012; Mulcahy et al., 2005) as well as capuchin monkeys (Anderson & Henneman, 1994) and baboons (Bolwig, 1963) can solve sequential tool use tasks spontaneously. Rooks, who have not been observed to use tools in the wild, also show spontaneous sequential tool use in experiments (Bird & Emery, 2009).

3.1.3 Associative tool-use in humans

In our view, ATU has received more attention from animal researchers than from those studying humans. An exception may be archaeologists studying complex tool composites and secondary tool use and their role in human evolution. These researchers have set out to investigate the cognitive processes involved in these types of ATU, stressing the role of executive functions, teaching, and language (Coolidge & Wynn, 2001; Haidle, 2010; Lombard & Haidle, 2012; Morgan et al., 2015).They also aim to determine when in phylogeny humans became capable of using tools to make tools and whether this ability represents a defining feature of the Homo lineage or whether it was shared by more ancient ancestors (Haidle et al., 2015; Harmand et al., 2015; J. W. K. Harris, 1983; Kitahara-Frisch, 1993; Schick & Toth, 2000).

Nevertheless, research into the *ontogenetic* development of ATU has been sparse. This may be because the “simpler forms” of ATU (i.e., sequential tool use, tool sets, metatool use rather than complex tool composites such as the bow and arrow) might appear trivially easy from a human adult perspective and so might not appear as an interesting study object. Indeed, as Köhler (1925) pointed out, most tool-use behaviours in adults have become “mechanized” (p. 65), i.e., are carried out with ease and so may obscure the ontogenetic and phylogenetic origins of these behaviours. Developmental studies on ATU in human children are needed as they can answer whether some forms of ATU can be acquired without social learning and whether the ATU behaviours that appear easy to adults can also be easily acquired by children (see introductory quote to this chapter). To our knowledge, there are only three studies on children’s ATU, all of which focus on sequential tool use: Alpert (1928) and Matheson (1931) replicated Köhler’s (1925) famous tool-use tasks in captive chimpanzees with preschool children. In their studies, 2- to 4-year-olds were separated from a reward by a railing set up in the testing room. Participants were required to use a short stick lying on their side of the railing to rake in a longer out-of-reach stick lying beyond the railing, and they could then use the long stick to obtain the reward. Results showed that it was not until 3 years of age that children solved this task and in general success rates were low: only 40% of the 3- to 4-year-olds in Matheson's (1931) study were able to solve the task. In a more recent study by Metevier (2006), 3-year-olds sat at a table and completed two tasks involving the use of a single tool before being presented with a sequential tool use task: In the "tube task", subjects had to use a stick to push a toy out of a tube. In the "rake task", they had to use a rake to obtain an out-of-reach toy on the table. Children found these tasks rather easy, indicated by success rates above 75%. Next, children were presented with a 'combination task', in which they first had to use the rake to obtain the stick lying out of reach on the table, and subsequently use the stick to push the toy out of the tube. Success rates were low – ranging between 25% and 37% – indicating that 3-year-olds struggled with the sequential tool use task even though they readily solved the individual components beforehand. In sum, Alpert's (1928) and Matheson's (1931) studies show that young children are able to solve sequential tool use tasks on their own, but that this type of ATU is challenging for them. Metevier's (2006) study shows that sequential tool use remains difficult even when children have received previous experience with individual elements of the tasks.

3.1.4 Research question

Köhler stated in 1925 that researchers were facing terra incognita with regard to understanding children's flexible tool behaviours – and not much seems to have changed. Similarly, Alex Kacelnik (pers. comm.) noted that only little is known about the developmental processes taking place between being a naïve newborn (human or other animal) and becoming a proficient tool user. Thus, the aim of the current study was to shed more light onto the developmental origins of different ATU types in human children.

Specifically, we studied young children's ability to spontaneously engage in three types of ATU: tool set, metatool use, and sequential tool use. For this, we designed six tasks: two tasks for each ATU type to be able to tease apart task-specific effects from effects resulting from the ATU type. Four tasks were based on spontaneous ATU behaviours in wild or captive animals (one tool set task, two metatool tasks, one sequential tool use task) and two were new creations. Each child was administered three tasks, with one task from each ATU category (task order counterbalanced). Similar to the logic in chapter 2, we investigated whether children were able to solve these tasks unaided. In addition, we examined whether children's success rates differed between ATU types.

3.2 Methods

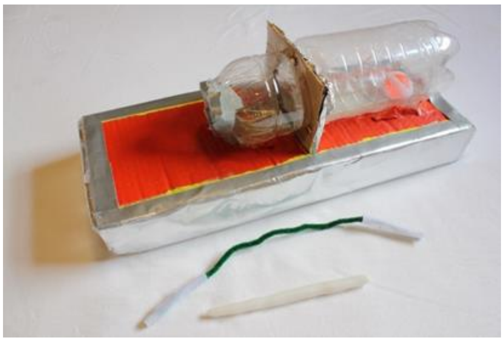
3.2.1 Participants

Sixty-six children (31 boys) between 3 years 6 months and 4 years 9 months (dates of birth were known for 64 children: M_{age} = 4 years 1 month (*SD* = 3.88 months)) were tested in nurseries and a Science museum in Birmingham, UK. The ethnic composition of the sample was 65.2% Caucasian, 21.2% Black, and 13.6% Asian. Participants tested in nurseries were recruited through letters sent to parents after an initial contact with the nurseries had been made; children tested at the museum were recruited via advertisements on the museum website. Children were rewarded with stickers regardless of success.

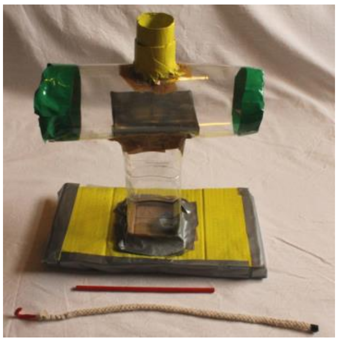
3.2.2 Material

The apparatuses are presented in Fig. 7 and described below. Tasks were designed so that they could be solved efficiently with a tool. For all tasks, tools were positioned between the apparatus(es) and the participant.

Tool set tasks



Open and probe



T-Tube

Metatool tasks

Anvil prop

Sponge push-pull

Sequential tool use tasks

Stick stick

Stick stone

Fig. 7. Apparatuses used in the associative tool use study.

3.2.2.1 Tool set

<u>Open and probe</u>

Wild chimpanzees use tool sets to open up bee hives, ant nests or termite mounds: They use a stout stick to *open* the hive/nest/mound and then use a thinner stick to *probe* for honey or insects (Shumaker et al., 2011). In our task, children had to retrieve a ball wrapped in Velcro from a transparent horizontal tube (l = 28 cm) by first inserting a plastic stick (l = 14.5) into the opening of the tube (diameter = 4 cm) in order to pierce a foil that was positioned in the middle of the tube, after which they had to use a pipecleaner (l = 29 cm) with Velcro strips attached to its ends in order to reach through the pierced foil into the rear of the bottle to retrieve the ball. The stick could only be used to pierce the foil and was too short to reach the ball. The pipecleaner was too bendy to be able to pierce the foil. The tube was placed in front of the child with the opening facing to the right.

<u>T-Tube</u>

This task was entirely novel. It consisted of a transparent T-shaped apparatus (a tube (l = 24 cm, diameter = 8 cm) attached horizontally to a vertical tube (7.5 cm x 7.5 cm x 13 cm)) and the goal was to retrieve a bucket containing a sticker from the bottom of the apparatus. Children had to first insert a stick (l = 19 cm) into either side of the horizontal part of the apparatus where there were two small openings (b = 7 cm, h = 2.3 cm) in order to displace a barrier (a rectangular piece of sponge covered with grey tape, 10 x 6.5 x 1.5 cm). Next, they had to insert a rope with a hook (l = 41.5 cm) into the top (diameter = 4 cm) of the apparatus to reach down to the bucket.

3.2.2.2 Metatool use

Anvil prop

Some wild chimpanzees engaged in nutcracking (i.e., using a stone hammer on a stone anvil to crack open nuts) have been observed to take one or even two other stones as wedges (metatools) to stabilise the surface of the anvil (Carvalho et al., 2008; Matsuzawa, 1991a, 1994). In our task, children had to break a plastic "nut" (diameter = 3.5 cm) on a wooden anvil (26 x 10 x 10 cm) using a wooden stick (l = 12.5 cm, diameter = 3 cm) as a hammer. Since the anvil was pyramid-shaped, it was always crooked, regardless of how it was positioned, and thus required stabilisation with a wedge (metatool; a piece of clay, ~ 10 x 6 x 6 cm). A blanket was placed in front of the children, preventing them from using the floor/table as a surface on which they could pound the nut. The anvil was placed on top of the blanket; the rest of the materials were between the anvil and the child.

Sponge push-pull

Wild chimpanzees have been observed to use a stick (metatool) to push a leaf further down into a tree-hole in order to retrieve water (Matsuzawa, 1991b, as cited in Sugiyama, 1997; Whiten et al., 1999). Similarly, captive orangutans have been observed to use a stick to push a paper towel into a puddle of sweet liquid (Shumaker et al., 2011). In our task, children had to use a wooden stick (l = 47.5 cm, diameter = 0.8 cm) to push a ball of cotton wool into a transparent tube (l = 24 cm, diameter = 7 cm) filled with 500 ml of water. The goal was to fill a small container placed next to the tube with water. The stick could be used on its own – dipping it into the tube to extract water – but this method was inefficient as it required the repeated use of the stick. Instead, the stick use could be improved by the wool as a metatool: The wool could be inserted into the tube so that it could absorb water, after which the stick could be used to retrieve the wool[14]. The wool had to be wrung just once to fill the small container.

3.2.2.3 Sequential tool use tasks

Stick stick

This task was based on an apparatus used by Taylor et al. (2007) studying sequential tool use in New Caledonian crows (see Fig. 1 in Taylor et al., 2007). The authors presented crows with a small box containing meat, a bigger box with vertical bars at the front that contained a long stick which was needed to retrieve the meat, and a short stick lying in front of the boxes. The birds had to use the short stick to rake the long stick closer to the bars of the big box to be able to retrieve the long stick. After that, they could insert the long stick into the small box to extract the meat (the short stick was too short). Three out of seven crows solved this task in the first trial. In our task, children had to use a short stick (l = 19.2 cm) to retrieve a longer stick (l = 31 cm) from a big box (also containing vertical bars at the front); then they could use the long stick to retrieve a bucket containing a sticker from a small box (30 x 21 x 5 cm) that was placed next to the big box. The small box had a narrow entrance (h = 1 cm) at the front where both tools could be inserted. However, the short stick was too short to reach the bucket, so it could only be used to obtain the longer stick.

Stick stone

This task was the second completely novel task. Metevier (2006) criticized that in previous sequential tool use studies the tool types were the same (e.g., the first and second

[14] Note that in our task the metatool was the wool (improving the stick), while in the cases of the apes the stick was the metatool (enhancing the reach of the paper/leaf).

tool were both sticks) and had to be used in the same way (e.g., for raking; see also our Stick stick task), which might have the made the tasks too easy to be representative for children's ATU abilities. Therefore, for the current task, we chose two different kinds of tools (a stick and a stone) that had to be used in different ways (poking, dropping): Children were presented with two apparatuses and a thin wooden stick (l = 18.5 cm, diameter = 0.4 cm). The task was to insert the stick into the smaller of the two apparatuses in order to push out a small stone. This stone could then be dropped into the top of the second apparatus where it would activate a trapdoor and release a pom pom. If the pom pom was released, children received a sticker. The apparatus containing the stone was a transparent L-shaped tube (19 x 6.5 x 7 cm) that was mounted on a base (h = 10.5 cm) in a way that the L was turned 90° clockwise. The horizontal part of the tube had a narrow opening at its side through which the stick could be inserted to push the stone (lying in the middle of the horizontal part of the tube) towards the other end whose entrance was facing the floor. The second apparatus (28 x 20 x 45 cm) was a transparent box with a trapdoor inside held in place by magnets. The pom pom was lying on the trapdoor. Dropping the stone into the apparatus and onto the trapdoor would open the door and release the pom pom. The apparatuses were placed next to each other.

3.2.3 Procedure

Participants were tested individually by the same female E and were sitting at a table or on the floor; E sat perpendicular to the child. In a warm-up game children were familiarised with breaking material within the experimental session by breaking sticks from Balsa wood into smaller pieces (see chapter 2). For the ATU tasks, children were presented with three semi-randomly chosen tasks, one from each ATU category. Tasks were presented as a game to the children in which they could win sticker. For each task, the materials were placed in front of the child and children were told the goal of the task, e.g., “to get this orange ball out of the bottle” (Open and probe). Children were told that they could “use anything here on the table/floor”, but they were never told that they had to use the tools to solve the tasks and they only received general encouragement. Children had 3 min to solve each game. Trials ended when children obtained the target, when time was over or when children refused to play. When one trial ended, E cleared the table/floor and fetched the next task. Children were never shown the correct solution of a task. Total testing time was ~ 15 min.

3.2.4 Scoring and analysis

Children’s behaviour was live-coded (video data were only available for nine children). We scored several binary DVs, two of which were used for the current analysis. *Associative tool use* measured whether children used both tools to solve the task in a manner that was intended by us. *Correct success* scored whether children succeeded after having used both tools in the correct way. We also scored *incorrect success* which indicated whether children solved the task in a way that was not intended by us. In order to obtain inter-observer reliability, 31% of the valid trials (i.e., n = 50) were live-coded by a second rater who was asked to code our two main variables, associative tool use and correct success. Inter-rater agreement for both variables were perfect (k = 1.000).

Data analysis consisted of three steps: A descriptive analysis investigated children’s performance in the individual tasks and across the ATU types to see whether children would be able to spontaneously solve the different types of ATU. Second, we examined whether the tasks within each ATU category were comparable in difficulty by conducting Chi-square analyses of children's rates of associative tool use and correct success within each ATU group. Lastly, we examined whether ATU type had an effect on children's associative tool use and correct success using a GLMM in R version 3.2.3 (R Core Team 2015).

3.3 Results

Each of the 66 children participated in three tasks, resulting in 198 trials. From these, 12 trials had to be excluded due to experimenter error (*n* = 7), failure of the apparatus (*n* = 4) or because the child became upset (*n* = 1). We also excluded all Anvil prop trials (*n* = 28) as we judged the task design to have failed: No child scored associative tool use nor correct success, whereas 71% of the participants were able to open the plastic nut in a way not intended by the task design (incorrect success): 11 children held the nut in one hand and directly hit it on the anvil; nine participants positioned the nut on the anvil and held it with one hand while using either the stick (*n* = 7) or the clay ball (*n* = 2) as a hammer. Since alternative ways of solving the Anvil prop task turned out to be readily available to the children, we judged the lack of associative tool use in this task to not represent a true lack of children's ATU capacities. Thus, this task was unable to answer the question whether children would be able to solve a metatool use task, and so we excluded it from further analyses, resulting in a final number of 158 valid trials across five tasks.

Figure 8 depicts children's rates of associative tool use and correct success for the individual tasks and across ATU groups. Overall, these rates were low. However, this was not a problem as we did not expect the majority of children to solve these tasks – rather, we were testing whether young children were *in principle* able to solve the tasks (similar to the logic in chapter 2). This was indeed the case, with two or more children in each task showing associative tool use and correct success. Open and probe revealed the highest percentages in both DVs, and it was the only task in which percentages were above 50%. This suggests that Open and probe was the easiest among our ATU tasks. The hardest tasks appeared to be the sequential tool use tasks in which only two children each scored associative tool use and correct success.

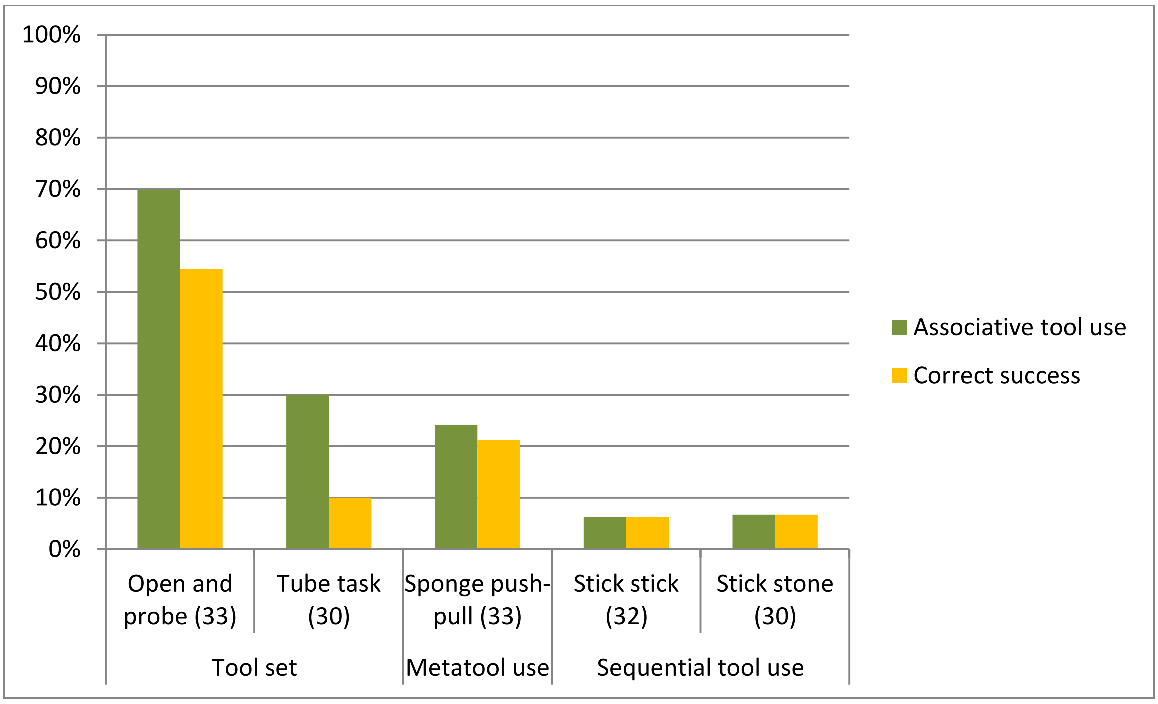


Fig. 8. Rates of associative tool use and correct success for the ATU tasks. Numbers in brackets are the sample sizes.

In order to examine whether the two tasks within the tool set and the sequential tool use groups were comparable in difficulty, we conducted Chi-square analyses. Results showed that for tool set, Open and probe had significantly higher rates of associative tool use ($\chi^2(1) = 9.908$, $p = .002$) and correct success ($\chi^2(1) = 14.032$, $p < .001$) than the Tube task, indicating that the tool set tasks were not of equal difficulty. With regard to the sequential tool use tasks, we found that Stick stick and Stick stone did not differ significantly in their associative tool use (Fisher's exact test, $p = .082$) nor correct success rates (Fisher's exact test, $p = .149$).

To investigate whether the ATU types differed in difficulty we conducted two GLMMs with binomial response variables using the glmer function of the R package lme4 (D. Bates et al., 2013). We created one model for each DV (associative tool use, correct success). In both models, sex, age in months (z-transformed to a mean of zero and a standard deviation of 1)[15], and ATU group (tool set, metatool use, sequential tool use) were included as independent variables (IVs); subject was included as a random effect to account for the fact that children contributed several datapoints. We checked the model assumptions and found that collinearity was no issue (largest VIF = 1.003 for associative tool use model; 1.001 for correct success model). For each DV, we assessed model stability by repeatedly comparing the model with models based on a reduced data set (excluding one case each time) and found no influential cases.

Results showed that ATU type had a significant effect on children's rates of associative tool use, $\chi 2\ (2) = 33.69$, $p < .001$ (sex and age had no significant effects and are not reported here to conserve space). Post-hoc Tukey tests conducted with the glht function of the R package multcomp (Hothorn, Bretz, & Westfall, 2008) showed that tool set was significantly easier than both metatool use ($p = .046$) and sequential tool use ($p < .001$), and that metatool use was significantly easier than sequential tool use ($p = 0.050$): Children were 3.3 (95% CI [1.02; 10.87]) times more likely to show associative tool use in the tool

[15] The missing age information for two of the children was replaced by the mean (49 months).

set tasks than in the metatool use task and 16 (95% CI [4.69; 54.33]) times more likely than in the sequential tool use task. Children were 4.8 (95% CI [1.29; 17.88]) times more likely to show associative tool use in the metatool task than in the sequential tool use task. However, the CI in the latter two comparisons (comparing the sequential tool use tasks with the other two ATU types) are quite large, indicating the presence of some uncertainty in the model. This might be due to the fact that only four children scored associative tool use in the sequential tool use tasks (small cell size) and so these results need to be treated with caution.

With regard to correct success, we found that ATU type had a significant effect on children's probability to succeed, $\chi 2$ (2) = 16.61, $p < .001$ (sex and age had no significant effects). Post-hoc Tukey tests showed that tool set tasks were significantly easier than sequential tool use tasks ($p = .002$): Children were 9.42 (95% CI [2.57; 34.53]) times more likely to show correct success in the tool set tasks than in the sequential tool use tasks (again note the large CI).

3.4 Discussion

This study investigated whether 3- and 4-year-olds were able to independently invent to use two tools in different combinations to solve several problem-solving tasks. To my knowledge, it is the first study on human children that investigates several ATU types at once. As Köhler (1925) already noted, the literature on (associative) tool use needs to be extended by studies on humans in order to provide a human comparison for the findings in animals. For this, it is also important to use tasks that are comparable across species. Therefore, we based our tasks – where possible – on behaviours observed in wild or captive animals (four out of six tasks) to increase the ecological validity of our tasks and to allow for species comparisons.

Children succeeded in all three types of ATU individually, without the need for social learning: tool set (using two tools to achieve a single outcome), metatool use (using one tool to improve a second tool), and sequential tool use (using a tool to get a second tool). We minimized the possibility that children could draw on directly relevant cultural knowledge when solving our tasks by testing children as young as possible and by presenting them with novel games that they were unlikely to have encountered before. However, we acknowledge that children did, and potentially also had to, draw on more *general* previous knowledge to solve the tasks, e.g., knowledge about affordances and physical properties of different materials (e.g., ropes or hooks). Assuming that we effectively limited children's ability to use *specific* cultural knowledge to solve the tasks, we conclude from our finding that children solved all tested ATU types that using two tools in combination potentially lies within the human ZLS, i.e., that at least some forms of ATU can be invented without social learning. Therefore, the current study adds to the tool behaviours identified as human latent solutions in chapter 2 some simple versions of tool sets, metatool use, and sequential tool use.

Children in the current study were older (3.5 to 5 years) than those tested for the study presented in chapter 2 (2 to 3.5 years). This is because the current study investigated a more flexible and cognitively more demanding form of tool use. Thus, we had to test older children since they possess more advanced working memory capacities, planning abilities, and fine motor skills, i.e., general cognitive skills that are not specifically related to tool-use skills. The low success rates in the current study further indicate that children of an even younger age might not have been able to solve our tasks. Similarly, the sequential tool use study by Metevier (2006) showed that 3-year-olds did not perform well even though they had previous experience with the different steps of the task. This highlights the general need to adjust the age range studied in human latent solution tests to the general cognitive and/or motor demands of the tasks. However, note that with increasing age the *cultural* knowledge that participants bring to the task also increases, making it more difficult to ensure that participants invent the correct solutions through *individual* (trial-and-error) learning. This points to a general challenge when creating latent solutions tests for humans, in that researchers need to find a trade-off between participants' young age and novelty of the tasks on the one hand and ensuring that participants already possess the domain-general cognitive abilities (e.g., executive functions, working memory) and motor skills to potentially solve the tasks on their own.

Success rates in our study were generally low, with only one task (Open and probe (tool set)) having a success rate of more than 50%, suggesting that the individual invention of relatively complex tool use is challenging but not impossible for young children. This highlights the point made by Köhler (1925) that even though the studied tool-use behaviours appear easy, if not "primitive" (p. 65), to human adults, this does not mean that the acquisition of these behaviours is also easy. Studies on children like the current one are needed to explore the basis of human tool use, demonstrating which tool behaviours are within children's spontaneous reach, upon which culturally acquired more complex tool-use skills can be added. In addition, researchers need to study children of a broad age range in order to investigate when children become able to solve certain tool-use tasks on their own.

We found that sequential tool use tasks were especially challenging for children. This is likely due the fact that the tasks involved two apparatuses, which might have increased the general task difficulty compared to the other ATU types. However, note that despite this, our tasks were still relatively easy versions of sequential tool use: We presented only one initial tool (rather than a choice of tools), with which only one other tool could be retrieved; the cost for retrieving the second tool was relatively small and the two apparatuses were in close proximity (in contrast, in some bird studies, apparatuses are positioned opposite of each other so that subjects need to turn around and thus keep the necessary information in their short-term memory, Wimpenny et al., 2009). The use of such design features in studies with great apes and birds has been criticised as making the tasks too easy and potentially overestimating the subjects' abilities (Martin-Ordas et al., 2012; Wimpenny et al., 2009), and in our case this means that we used a simple version of sequential tool use in order to detect any spontaneous abilities in young children. Nevertheless, we found the tasks to be very difficult for our participants. Future studies looking more closely at the development of ATU abilities could identify which cognitive mechanisms are involved in sequential tool use and what makes it so challenging for young children.

We also found that the different types of ATU were not equally easy, with tool set being the easiest type of ATU, followed by metatool use, and then sequential tool use as the hardest type. This could be interpreted as a first hint at a potential "cognition-based hierarchical organization" (Shumaker et al., 2011, p. 21) of these ATU types. Yet, our results should be treated with caution as long as there is no replication of the findings. This is because first of all, this is – to our knowledge – the first study investigating different ATU types in humans and thus finding such a result. Second, as mentioned above, the poor performance in the sequential tool use tasks might be explained by the fact that the tasks consisted of two apparatuses, making them less comparable to the tool set and metatool tasks. In addition, children's good performance in the tool set tasks might have been driven by Open and probe, and so future studies need to investigate whether this result is due to a task-specific effect or whether tool set tasks are generally easier than other ATU types. Finally, due to the exclusion of Anvil prop, the metatool use category was only represented by a single task (Sponge push-pull) and so an effect based on this ATU category cannot be distinguished from an effect specific to the Sponge push-pull task. More studies are needed to investigate whether there is a hierarchy of different ATU types. So far, it is assumed that there is no hierarchy, but this is mainly due to absence of evidence rather than to evidence of its absence (Shumaker et al., 2011).

Our study was exploratory, looking at whether young children can spontaneously invent different ATU types. Due to this quite basic question and the use of a binary outcome variables (success yes/no), the data were condensed substantially for the current analyses. However, our data are much richer and could also be analysed with a focus on the process of how children (individually) learned to solve the tasks, which steps proved difficult and where children got stuck, why children stopped and/or failed, to what degree chance played a role for success or what kind of mistakes children made (for example, Köhler, 1925, suggested differentiating between "good" and "bad" mistakes). These are important questions which can give us insights into which cognitive mechanisms underlie the acquisition and application of flexible tool use. With these questions, the study of tool use also dovetails with the study of general problem-solving abilities (see the approach by Köhler, 1925, for chimpanzees and by Alpert, 1928, and Matheson, 1931, for human children), and indeed, the study of tool use is regarded as an excellent means to investigate problem-solving (see also Cutting, Apperly, Chappell, & Beck, 2011; Keen, 2011). However, investigating these questions (i.e., focusing on the process of tool using rather than its binary outcomes (success yes/no)) is a different research focus and is beyond the scope of this thesis. Finally, future studies could also look at individual differences in children's performance to identify cognitive and motivational factors accompanying the development of ATU abilities (e.g., executive functions, impulsivity, handedness).

3.5 Interim summary and link to chapter 4

Köhler (1925) noted that there is the need to conduct experiments investigating which cognitive skills and behaviours a given species can achieve on its own (as he himself did in his studies with chimpanzees and, in the case of humans, e.g., Alpert, 1928, Matheson, 1931, Piaget, 1952, or E. Bates et al., 1980). By providing evidence on which behaviours a species can acquire unaided this kind of research would form the necessary empirical basis for studies investigating what a species can achieve when social learning is available. Chapters 2 and 3 of this thesis have contributed to the literature in this way by investigating what kinds of (associative) tool-use behaviours humans can achieve individually, without cultural knowledge, by testing young children with novel problem-solving tasks.

Social learning studies, in turn, are important and have been the subject of much research efforts because social learning by definition lies at the heart of cultural evolution. Studies on social learning can grant us insights into the origins and characteristics of (human) culture and into the differences and similarities across species. The literature on social learning is vast (for a review, see e.g., Hoppitt & Laland, 2013), with some major insights from developmental psychology being that humans use a portfolio of social learning mechanisms from early on (Whiten et al., 2009), that children are "cultural magnets" (Flynn, 2008), readily and often actively absorbing cultural knowledge (see also Chouinard, 2007), that they exhibit a range of social learning biases (Wood, Kendal, & Flynn, 2013) and that they also imitate for purely social reasons (Over & Carpenter, 2012; Uzgiris, 1981). We also know that faithful social learning has allowed our species to produce cumulative culture (Boyd & Richerson, 1996; Tomasello, 1999a) and that – on an individual level – social learning allows humans to acquire traits that they could not have invented on their own (culture-dependent traits, chapter 1; see also Vygotsky, 1978).

For example, as we will show in the introduction of the following chapter, children from their second year of life are able to learn novel social-conventional acts – behaviours that are highly arbitrary and thus unlikely to be invented spontaneously (e.g., switching on a light with the forehead rather than with the hand, Gergely, Bekkering, & Király, 2002). However, there is a research gap with regard to the material culture (as opposed to the social-conventional culture) domain: To our knowledge, prior to this thesis it had not been investigated at what age children start to acquire culture-dependent traits in a material cultural context. The next chapter presents two studies in which we start filling this knowledge gap. Thus, we will now leave the latent solution test approach aside and I will – in the remainder of the empirical part of this thesis – present our social learning studies.

CHAPTER 4: YOUNG CHILDREN COPY CUMULATIVE TECHNOLOGICAL DESIGN IN THE ABSENCE OF ACTION INFORMATION

The following chapter is a modified version of the paper published as:

Reindl, E., Apperly, I. A., Beck, S. R., & Tennie, C. (2017). Young children copy cumulative technological design in the absence of action information. *Scientific Reports, 7*(1), 1788. doi: 10.1038/s41598-017-01715-2

For this chapter, the main text and the supplementary material of the paper have been rearranged to allow for better readability. Minor modifications have been made to the introduction, methods, results, discussion, but otherwise the text is as published.

I am the primary author of this publication. The original idea for this study was developed in collaboration with my supervisors Claudio Tennie, Sarah Beck, and Ian Apperly. I was primarily responsible for the design of the studies and carried out all data collection and analysis. My supervisors contributed to authorship by providing feedback and editing versions of this paper leading to its publication.

Link to open access article: https://www.nature.com/articles/s41598-017-01715-2

4.1 General introduction

Humans' capacity to spread across the planet and to reach out beyond its boundaries has often been explained by our ability to produce cumulative culture, i.e., to accumulate changes in cultural traits beyond a level that individuals can reach on their own (Boyd et al., 2011; Henrich, 2015). These changes entail both improvements and deterioration as well as changes that have no effect on trait efficiency, but it is our capacity to accumulate *beneficial* modifications over time – a phenomenon labelled the ratchet effect (Tennie et al., 2009; Tomasello, 1999a) – that is thought to be among the key characteristics setting us apart from other animals, including culture-bearing species such as chimpanzees and orangutans (Tomasello, 1999a). Identifying the cognitive processes underpinning the ratchet effect will help us understand more about how human culture has evolved.

Researchers have begun to experimentally investigate the cognitive mechanisms involved in cumulative culture: Methods have been developed to simulate cumulative cultural evolution in the laboratory, e.g., behavioural experiments using the transmission chain paradigm (Caldwell & Millen, 2008a, 2009; Caldwell, Schillinger, Evans, & Hopper, 2012; Caldwell & Smith, 2012; Morgan et al., 2015; Osiurak et al., 2016; Wasielewski, 2014; Zwirner & Thornton, 2015) or virtual tasks investigating the effects of social learning in groups (Derex & Boyd, 2015; Derex, Godelle, & Raymond, 2012). These studies show that human adults readily exhibit a ratchet effect. In conjunction with other experimental studies (Dean et al., 2012; Schillinger, Mesoudi, & Lycett, 2015) and modelling approaches (Acerbi, Jacquet, & Tennie, 2012; Lewis & Laland, 2012) these findings suggest that the capacity for high-fidelity social transmission is a crucial prerequisite for the ratchet effect – alongside a capacity for innovations (Enquist et al., 2008) and other relevant cognitive mechanisms (Dean et al., 2014).

Which social learning mechanisms enable the production of a ratchet effect? Many social learning mechanisms such as stimulus or local enhancement are capable of supporting culture over even extended time, yet they have been argued to be of insufficient fidelity and bandwidth to *accumulate* culture (Boyd & Richerson, 1996; Galef, 2012; see also ZLS-only account, Reindl et al., in press, Tennie et al., 2009). In these cases, the learning mechanisms draw the learner's attention towards a stimulus or location, after which the behaviour in question is acquired through individual learning. Thus, learning is rather an individual response than a faithful copy. This also applies to instances in which learners encounter cultural artefacts that were left behind by conspecifics (*exposure*, e.g., used tools), even where these artefacts "canalize" the learner's behaviour in important ways (Tennie et al., 2009; Reindl et al., in press)[16]. As a consequence, the range of traits that a learner can acquire by this combination of low-fidelity social learning and individual learning is limited to those that the learner can actually invent individually (i.e., to latent

[16] An interesting question is whether exposure to *cumulative* cultural artefacts might be able to sustain a ratchet effect (e.g., imagine a scenario in which there was no overlap between generations, so that each new generation would only have access to the products left behind by the previous generation; or imagine a group of early hominins that encounters a camp of another group an finds novel artefacts). We would argue that while encountering novel cumulative cultural products can transfer valuable information to a naïve learner and possibly enable her to make use of the artefact, thus letting her go beyond what she could have achieved individually, many cumulative cultural products would be too complex and opaque to provide much benefit when encountered out of context. It might often be impossible to find out how, when and for which purpose the artefact in question would have been used for. Of course, these questions need rigorous scientific investigation to be answered eventually.

solutions, Tennie et al., 2009). Traits that are too complex or unlikely to be invented individually (culture-dependent traits) – a defining feature of cumulative culture (Boyd & Richerson, 1996) – cannot be acquired by mechanisms that only harness the power of individual learning (such as stimulus and local enhancement). Instead, cumulative cultural traits need high-fidelity copying, e.g., copying of the actions and/or end-results of a trait, in order to be acquired (but see Henrich & Boyd, 2002; Morin, 2016a).

It has been argued, and demonstrated, that cumulative culture can be transmitted via imitation (Boyd et al., 2011; Derex & Boyd, 2015; Derex et al., 2012; Osiurak et al., 2016; Schillinger et al., 2015; Tennie et al., 2009, 2012; Tomasello, 1999a; Wasielewski, 2014).Whether emulation, i.e., learning about effects or results in the environment (Carpenter & Call, 2002), can ever be sufficiently faithful to sustain a ratchet effect is still debated (Caldwell & Millen, 2009; Caldwell et al., 2012; Galef, 2012; Schillinger et al., 2015; Tomasello, 1999b; Zwirner & Thornton, 2015).

The general capacity for high-fidelity social learning is within the human cognitive repertoire from an early age (Want & Harris, 2002): Before the end of their first year, infants are able to copy novel actions (R. Barr, Dowden, & Hayne, 1996); by 1 year, they flexibly switch between emulation and imitation (McGuigan & Whiten, 2009; Nielsen, 2006) and by the end of their second year, children imitate even causally irrelevant actions (*overimitation*; Kenward, 2012; Keupp, Behne, & Rakoczy, 2013; Lyons, Young, & Keil, 2007; Marsh, Ropar, & Hamilton, 2014; McGuigan, 2013; McGuigan & Robertson, 2015; Nielsen & Blank, 2011) – a propensity argued to be highly adaptive in an environment of cognitively opaque cultural artefacts and skills (Gergely & Csibra, 2006). Given the special role high-fidelity copying plays for cumulative culture, these findings indicate that young children already possess some crucial cognitive prerequisites for acquiring and transmitting cumulative culture.

Preceding these findings, Vygotsky (1978) theorized that children's ability to imitate others enables them to acquire knowledge and skills that they cannot (yet) learn individually but which can only be acquired via social learning and teaching: Imitation and teaching allows humans to extend their current behavioural repertoire (the ZAD) by novel skills and knowledge that require social learning to be acquired (what later was labelled cumulative culture, see chapter 1). Vygotsky also pointed out that individuals "can imitate only that which is within [their] developmental level" (ibid., p. 88), i.e., within their ZPD. For example, a child cannot learn higher mathematics until she has mastered more basic concepts such as adding, subtracting or understanding the concept of zero. Thus, the range of cumulative cultural traits that can be acquired by a given individual at a given point in development is not infinite but is limited by the individual's ZPD whose content and range depends on the individual's ZAD.

So far, no study has investigated children's capacity to learn from or copy traits that they could not have invented on their own in the technological domain. This is surprising as the acquisition of technological skills is an important form of cultural transmission in our species, responsible for the vast amount and complexity of technology accumulated today. This chapter examined children's capacity for copying a "culture-dependent trait in the technological domain", which we label *cumulative technological design*. We define *technological design* as a material cultural product created by a sequence of instrumental actions, i.e., actions which "bring about a tangible, functional outcome" (Legare & Nielsen, 2015, p. 692) and which are causally – i.e., non-arbitrarily – linked to this outcome. The production of technological design is a subtype of *instrumental skills* (Legare

& Nielsen, 2015): Instrumental skills are those that achieve functional outcomes by either arbitrary (e.g., typing in a number combination to unlock a phone) or non-arbitrary (e.g., levering to open a box) actions (with arbitrary actions requiring the learner to pay relatively more attention to the actions compared to the results of the demonstration). Technological design refers to products that are created by non-arbitrary actions only, and are thus inherently more results-focused.

Previous studies on children's social learning often aimed to differentiate between imitative and emulative learning (Flynn & Whiten, 2010; Want & Harris, 2002), and thus tested children's motivation and ability to act on and/or manipulate parts of the environment based on demonstrations. None of these studies tested for the recreation of technological design. Nevertheless, they provided important insights into children's copying behaviour, many of which are relevant to understanding the acquisition of (cumulative) technological design. The studies that come closest to testing whether children can copy something that they cannot invent on their own are those that investigated children's ability to copy social conventional acts (Legare & Nielsen, 2015). Social conventional acts are usually those in which the relationship between the outcome and (parts of) the actions is not causal, but arbitrary (thus conventional), e.g., driving on the left or the right side of the road or tapping the side of a box before opening it (Horner & Whiten, 2005; Lyons et al., 2007; Marsh et al., 2014, McGuigan & Graham, 2010; McGuigan & Whiten, 2009). In addition, acts can be interpreted as social conventional if they contain obviously inefficient actions, e.g., switching on a light with the forehead rather than with the hand (Gergely et al., 2002; Nielsen, 2006) or if the start states and end states of these acts do not differ (Watson-Jones, Legare, Whitehouse, & Clegg, 2014). These studies show that from their second year of life children are able to copy novel conventional acts. As some of these behaviours were not shown spontaneously by the children (e.g., turning on a light with the head), these behaviours likely represent culture-dependent behaviours for children in the respective age ranges.

While these studies tested for children's ability to copy social conventional cumulative culture, the question of children's capacity for copying technological cumulative culture (cumulative technological design) remains unanswered. Conclusions drawn from studies involving social conventional actions may not apply to the technological domain, not least because in the latter both imitation (copying goals, results, and actions) and emulation (copying goals and results, but using different actions) may be important (Acerbi & Tennie, 2016; Heyes, 2013). For example, technological demonstrations allow the learner to gain information by attending to endstates as well as intermediate states. Therefore, in contrast to social conventional demonstrations, learners could focus only on the outcome and then reproduce it using their own means (emulation, Carpenter & Call, 2002). Technological tasks also allow learners to combine action and results copying, resulting in greater transmission fidelity than that which copying a single type of information alone would be able to achieve (*redundant copying*, Acerbi & Tennie, 2016). Thus, social conventional and technological demonstrations differ systematically with regard to the social learning mechanisms that can be involved, which likely has implications for the transmission of cumulative culture.

Young children use imitation and emulation to perform instrumental tasks even before their first birthday (Flynn & Whiten, 2010; Hopper, Flynn, Wood, & Whiten, 2010; Nielsen, 2006; Thompson & Russell, 2004). However, the target actions or results in previous studies have only consisted of simple actions or results or combinations of those that participants in baseline conditions were able to invent on their own; i.e., these behaviours were likely within the respective study participants' ZAD. This chapter presents two studies that are the first to look at children's ability to copy cumulative technological design, i.e., technological design that children are not yet able to invent on their own.

We adapted a task that was used to study cumulative cultural evolution in adults: the "spaghetti tower task" (Caldwell & Millen, 2008a). In a first step, we needed to create a specimen of cumulative technological design, i.e., a technological artefact that children at a given age could not invent on their own. Therefore, in study 1 we 1) explored which age range the tower task was most suitable for, 2) investigated what children in this age range could achieve on their own (baseline performance), and 3) created a cumulative technological design that children in the baseline were not able to invent on their own. Using this cumulative technological design product, study 1 also investigated whether children who observed a demonstrator build the cumulative technological design would be able to learn from or even copy it. In a second study, we introduced amendments to the study design, replicated the results from study 1, and added another social learning condition (an endstate condition, requiring emulation learning).

4.2 Cumulative culture study 1

4.2.1 Introduction

Study 1 investigated whether young children already possess one of the key cognitive ingredients for cumulative cultural evolution: the ability to copy a novel trait which they could not have invented by themselves. For this, we first identified and then provided children with a cumulative technological product that was beyond what children could produce on their own (thus being a culture-dependent product) to see whether children would copy it and/or benefit from this demonstration in a way that allowed them to go beyond what they could have achieved on their own. We adapted Caldwell and Millen's (2008a) spaghetti tower task in which adults are asked to build something very tall from raw spaghetti and modelling clay (in our version, children use plasticine and plastic sticks). We chose this task for the same reasons mentioned by Caldwell and Millen: The task has a well-defined goal (building something as tall as possible) and allows a continuous and objective measurement of performance (height). In addition, child participants bring relatively few preconceived ideas to the test session. The task can also easily be implemented in schools for it uses only little space and time.

Children were tested in one of two conditions: The baseline group was simply asked to build something very tall. This condition informed us about the height and types of constructions children could reach without help, so data for this group were collected first. Children in the full demonstration group received the same instruction, but before they started, they watched the experimenter build a tower of a height and complexity that they were not able to achieve on their own. In order to create such a culture-dependent product for young children, we created a tower whose design and height did not occur among baseline children.

In cumulative cultural evolution, changes to a trait can occur in two domains (Dean et al., 2014; Tennie, Caldwell, & Dean, 2017): In its *complexity/design* (in our task tower *structure*) and in its *efficiency* (tower *height*)[17]. We focused on the design aspect and chose

[17] We acknowledge that for some readers the term "efficiency" might not seem intuitive for describing tower height. Other terms, such as "functionality" might seem more appropriate. However, we chose this labelling

– after we collected the data on the baseline children – as our cumulative technological design a *tripod*: a tower with a base of three legs arranged in a triangle (rather than, e.g., in a line) and combined at the top with a piece of plasticine (Fig. 8).

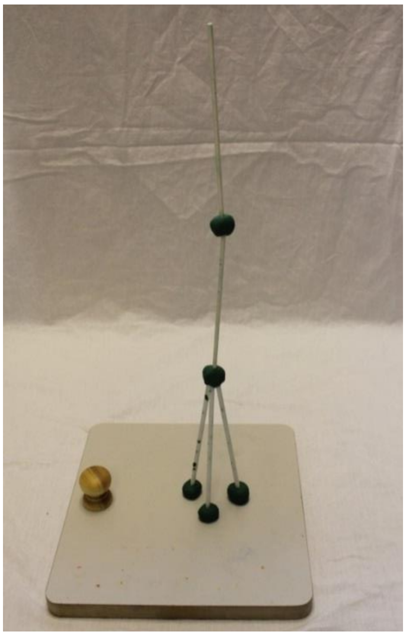


Fig. 9. Demonstration tower (tripod). Example of the cumulative technological design children in the full demonstration condition were presented with (h = 46 cm).

Cumulative culture is inherently open-ended (Tennie et al., 2017) and the tripod operationalizes this aspect by being a hierarchical and open-ended cultural product: The order of the building actions is determined through a hierarchy (e.g., "first construct the tower base, then build upwards. For the base, first form the plasticine, then insert the sticks, etc."); at the same time, the number of possible steps is not limited by the task (thus open-ended). We chose a hierarchical task because "most cultural products are compound products" (Wimsatt, 1999, p. 285), requiring a "lengthy sequence of actions […] with each

---

to relate to nomenclature that had previously been introduced to the literature ("efficiencies and complexities", Dean et al, 2014, p. 298).

action functionally dependent on previous actions” (O’Brien, Lyman, Mesoudi, & VanPool, 2010, p. 3801). Thus, we presented children with a *cultural recipe* (ibid.) rather than with simple actions or results or combinations of those as has been the case in previous studies (see above).

The tripod represents a superior design compared to other tower designs, as it allows for greater heights to be achieved. Evidence comes from studies using the spaghetti tower task with adults in which the tripod was one of the most efficient designs invented by participants (Caldwell & Eve, 2014; Caldwell et al., 2012). Moreover, Caldwell et al. (2012) found a positive relationship between the number of tripod design features in participants’ constructions and the height of these constructions – such a relationship was absent for a cubic design, suggesting that the tripod allows for greater heights to be achieved, and so comes closer to the notion of open-endedness. Finally, these studies also found that participants who were presented with tripod designs were themselves able to build constructions that were as tall as (Caldwell & Eve, 2014) or even significantly taller (Caldwell et al., 2012) than the demonstrated tripods, whereas participants presented with a different (cubic) design were not able to go beyond the demonstrated height. Note that the fact that the tripod is superior to many other designs does not imply that participants choosing the tripod design will automatically make towers that are taller than other shapes.

In our full demonstration condition we presented children with a tripod (both with the building actions and the end-result). This condition would inform us about children’s performance when given access to a culture-dependent product. Would they be able to use the information they gained from the demonstration to boost their performance beyond the scope of their individual inventiveness? Comparing the towers in the baseline and the full demonstration condition would provide us with first answers to this question.

4.2.2 Pilot study: Identifying an appropriate age range for the tower task

In a pilot study we determined the age window for which the tower task was most appropriate. Initial pilot work (not reported here) tested slightly different construction tasks (building a horizontal construction or a tower from materials such as Play Doh, wooden sticks or tape) and indicated that children below 4 years found any of these tasks very difficult. This was likely because of children's difficulties in understanding the goal of the game, their still developing understanding of the physics involved in the task (e.g., aspects related to gravity and mass), still developing fine motor abilities, or limited interest in and/or motivation to play the game. Therefore, we concluded that children below 4 years would likely be too young for our study.

In the pilot study, we tested for the first time our final tower task, with 15 children between 3 years 6 months and 4 years 4 months. Again, we found that several children in this age range tended to be too young for the task as they struggled with understanding the physical characteristics of the task, were limited in their fine motor abilities, or sometimes did not seem to understand the task or lacked enough interest. Therefore, we decided the lowest age for which the task was appropriate to be 4 years. We also collected data from three children aged 5 years 10 months. One child was assigned to the baseline condition and built a level-3-tower. Two children were tested in the full demonstration condition and showed good performance. We concluded from this pilot study that the tower task was most suitable for studies with children between 4 and 6 years of age.

4.2.2 Methods

4.2.2.1 Participants

We tested 34 children (15 boys) between 4 and 6 years (M_{age} = 4 years 10 months, *SD* = 5.14 months, range: 4 years 1 month to 5 years 9 months) in Birmingham, UK. The ethnic composition was 53% Caucasian, 29% Black, and 18% Asian. Baseline children (tested first) were recruited and tested in nursery schools. Children in the full demonstration condition (tested second) were recruited via advertisements on a local parenting website and on the website of the science museum where the testing for this condition took place. Parents willing for their child to participate gave written informed consent. Seventeen children (six boys) were assigned to the baseline, the other half to the full demonstration condition. There were no differences in age ($t(32) = 0.662$, $p = .513$, Cohen's $d = 0.228$) or sex ($\chi^2(1) = 0.477$, $p = .490$) between conditions. Participants were rewarded with a sticker regardless of success.

4.2.2.2 Material

Warm-up game

We used a day-night Stroop task consisting of 24 cards showing a picture of either daytime or nighttime (Gerstadt, Hong, & Diamond, 1994). Children should say "night" when shown the day card and "day" when shown the night card. We chose this game as pilot work showed that children were motivated to play this game, it required little space and material, and could be adapted to children's skill level. The task was used to familiarise children with E; responses were not recorded.

Construction task

Each participant was provided with 30 white solid plastic lollipop sticks (length = 15 cm, diameter = 4.5 mm) and 70 g of green plasticine, first presented as one ball and then formed into three separate balls during a short demonstration of the materials (see procedure). Children sat on the floor at a low table (40 x 35 x 15 cm). Constructions were built on a wooden board (25 x 25 x 2 cm) attached on top of the table with a bar clamp. At the end of each session, the clamp was opened and the board with the construction was transferred to another section of the testing area where photos were taken. Further materials were a folding rule to measure the height of the constructions, a stopwatch, a video and a still photo camera, and a white sheet as the photo background.

4.2.2.3 Procedure

Children were tested individually by E. After the warm-up, E told children that she had a special game for them which would be played at the table. She explained that the game was "to build something that is very high, as high as you can make it". E showed children the plastic sticks and the plasticine and said: "You can use these things to help you build it. You can do anything you like with these things to try to make something very tall. You can use all of these [E pointed to lollipop sticks] and all of this [pointed to plasticine]. Also, with this [E took plasticine] you can do things like this [E tore one third off of the ball] or this [E tore another third off, then rolled it]."

Participants in the baseline condition were told that they did "not have much time to build something that is as high as possible", so they needed to be quick. E placed the three plasticine balls on the table next to the board and the sticks on the floor in front of the children. Children were then encouraged to start building. Participants in the full demonstration condition were told that the game involved turn-taking, so that E would start first and that the child's task was to watch her. When the time was over, it would be the child's turn to "build something that is as high as possible". E then put the child's materials out of reach and fetched a new set to build a tripod. The construction was chosen to be 46 cm tall in order to be substantially greater than the maximum height achieved by the baseline group (33.5 cm). After building her tower (50-60 sec), E announced that it was the child's turn now. She put the board with the tripod on the floor to the left of the table and removed the rest of her building materials. Children were given a new board, three plasticine balls (lying next to the board), and 30 sticks (on the floor in front of the children). As in the baseline condition, children were told that they did "not have much time to build something that is as high as possible", so they needed to be quick. They were then encouraged to start building.

Building time in both conditions was 6 min. While children were building, E sat next to the table making notes and did not intervene. If participants did not begin building, E encouraged them by saying "Try to make something very high with these things!" If children asked for help, E replied "Let's see what you can do – see how high you can make it!" In cases where children said they were finished before the time was over, E encouraged them to continue by saying "You still have some time left! Try to make it even higher!"

When time was up, children were not allowed to touch the construction anymore. Children who held their construction in their hands were asked to place it on the table and those who stabilised it with their hands were told to let go. Towers that could not stand on their own had to be placed horizontally on the table. Tower height was measured at the tallest point of the construction (note that for towers that were lying horizontally, we measured the actual height, not the potential one when held upright). Once the participant left the room, E took pictures of the construction (one from each side, one from above). All constructions were destroyed before the next child arrived at the testing area and every participant received a new set of sticks and plasticine. Total testing time was 10 to 15 min.

4.2.2.4 Coding and analysis

With regard to the baseline condition, we were interested in the design and height of the towers that children were able to make on their own. Based on these data, we created the cumulative technological product for our study: the tripod. With regard to the full demonstration condition, we were interested in whether children would be able to copy the tripod from a full demonstration. In addition, we investigated whether children would build taller towers than children in the baseline condition, and how similar children's constructions were to the demonstrated tripod. For this, we measured three variables: *tower height* (measured at the end of the trial), *tower shape*, and *similarity to tripod* (similarity of the construction to the tripod). Analyses were carried out with IMB SPSS Statistics 22 unless indicated otherwise.

Tower height

Tower height represented the height of the construction at the end of the trial. Therefore, tower height did not necessarily reflect the maximum height achieved by the child, as in some cases towers crashed and could only partly be rebuilt in the remaining time. This was especially the case for children in the full demonstration condition: In 9 out of 17 cases, the towers with the maximum height crashed or were disassembled and rebuilt. Thus, at the end of the trial, towers with a smaller height than the maximum height achieved were measured, which potentially underestimated children's performance. In the baseline, there were only four children for whom the final tower did not represent the tower with the maximum height. Although we were not able to establish the exact height of towers that did not "survive" until the end of the trial, we were able to determine their height in "stick levels" (see results) from the videos.

To investigate whether final tower height differed between the baseline and full demonstration condition, we conducted an analysis of covariance (ANCOVA) with final tower height as DV, condition (baseline, full demo) as the IV, and age (in months) as a covariate. All assumptions for ANCOVA were met.

Tower shape

The shape of the tallest tower was coded offline by E based on photos and video stills. For children for whom the final towers were also the tallest towers, we based this classification on the photos taken at the end of the trial. For children for which the tallest tower was built throughout the session but did not "survive" until its end (because it crashed or because children disassembled it; baseline: n = 4, full demonstration: n = 9) we coded the shape of both the final tower (using the photos taken at the end of the trial) and of the tower with the maximum height (using stills from the video). One might ask why there were more instances of crashed towers in the demo compared to the baseline condition. One reason could be that in the demo condition, seeing a tall tower resulted in children not only making taller towers themselves (see results) but also less stable ones (as not all children copied the efficient tripod shape) and so towers in the demo condition were at higher risk of crashing.

First, we determined tower height in what we labelled *stick levels* (results: Table 7, first column): We counted how many sticks were vertically combined on top of each other ("combining" meaning two sticks joined vertically with a piece of plasticine, while an overlap of up to half the length of a stick was allowed). This allowed us to group the towers into four categories: *level-0-constructions* (towers that were smaller than the height of one stick, e.g., towers lying on their side or constructions consisting only of plasticine); *level-1-constructions* (constructions with one (or more) sticks placed vertically into a plasticine base); *level-2-constructions* (comprising any constructions in which two sticks were combined on top of each other); and – applying the same logic – *level-3*- and *level-4-constructions* (thus, the tripod fell into the level-3-constructions group). We then grouped the towers within each stick level category based on their shape, resulting in one to three shape categories per stick level (Table 7, columns 2 and 3).

<u>Similarity to tripod</u>

We established the similarity of the constructions (in both conditions) to the demonstrated tripod, using the same procedure as Caldwell and Millen (2008a): Two raters, blind to the research hypotheses, coded pictures of *all* towers: For children whose final tower was the tower with the maximum height, final towers were rated; for children whose final tower was not the one with the maximum height, both the final tower and the tallest tower were coded. Pictures were coded with regard to their similarity to the demonstrated tripod, using a scale from 1 (*not similar at all*) to 7 (*very similar)*. Points in between were not labelled. For each participant, raters were given one picture of the participant's construction, which was to be compared with a picture of the demonstration tripod (presented in five pictures, to show the slight, but unavoidable variance of the demonstrated tripods) and asked "How similar is this [image] to the constructions on the 5 pictures?" The ratings of the two coders correlated significantly (r = .565, p < .001); the strength of the relationship between the two ratings was similar to the one in Caldwell and Millen (2008a).

To determine whether the towers in the full demonstration condition were rated more similar to the tripod than the towers in the baseline, we fitted a linear mixed model (LMM, Baayen, 2008) in R (version 3.2.3; R Core Team, 2013) into which we included condition as a fixed effect and random intercepts for participant and rater (to account for the fact that each rater and some participants contributed more than one data point). We also included a random slope for condition (manually dummy-coded) on rater (D. J. Barr et al., 2013). The model was fitted using the function lmer of the R-package lme4 (D. Bates et al., 2013). To allow for a likelihood test, we fitted the model using Maximum Likelihood (rather than Restricted Maximum Likelihood, Bolker et al., 2008). We checked for normal distribution and homoscedasticity of the residuals by visually inspecting a qq-plot and the residuals plotted against the fitted values and found no obvious violation of these assumptions. The sample size for the model was 94 ratings made on 47 towers (two ratings per tower).

4.2.3 Results

4.2.3.1 Tower height

Across conditions, mean final tower height was 19.85 cm (*SD* = 11.68 cm), ranging from 3.5 to 44 cm. In the baseline, mean tower height was 15.56 cm (*SD* = 8.18 cm), ranging between 3.5 and 33.5 cm. In the full demonstration condition, mean tower height was 24.15 cm (*SD* = 13.24 cm), ranging from 4 to 44 cm. The ANCOVA revealed significant main effects for condition, $F(1,31) = 8.473$, $p = .007$, partial $\eta^2 = .215$, and age, $F(1,31) = 9.710$, $p = .004$, partial $\eta^2 = .239$, on tower height: Older children built significantly taller towers than younger children. On top of the age effect, children in the full demonstration condition built significantly taller towers than those in the baseline, suggesting that children benefitted from the demonstration of cumulative technological design.

The difference in tower height between conditions might have been even more pronounced had we measured children's towers not only at the end of the trial, but also throughout (as we did in study 2). Although we did not collect continuous data on tower height (in cm) throughout the trial, we recorded height in stick levels for every construction children made during the trial and were thus able to record maximum height in stick levels: In the baseline, the towers with the maximum height in stick levels were level-2-constructions; in the full demonstration condition, the tallest towers measured in stick levels were level-3-constructions. Since the variable stick levels was not normally distributed, we used Mann-Whitney U tests to determine whether there was a statistical difference between conditions with regard to both the final and the maximum tower height in stick levels. Results for the comparison of the *final* towers showed a statistically non-significant trend of final towers in the full demonstration condition reaching higher levels that those in baseline (U = 89.00, p = .057, Cohen's d = 0.795). This non-significant result might not be surprising as the difference in tower height *in cm* between the conditions was also not very pronounced (see above) and because the measurement in stick levels uses a coarser measurement scale. However, as suggested, the comparison of tower height of the *maximum* towers was more pronounced: Towers in the full demonstration condition were significantly taller (measured in stick levels) than those in the baseline (U = 53.50, p = .001, Cohen's d = 1.357).

4.2.3.2 Tower shape

No child in the baseline made a tripod (Table 7). The most common tower shapes in the baseline were hedgehogs and other level-2-constructions (both shapes were made by 4 out of 17 children). In the full demonstration condition, the most common tower shape was a level-2-tower (6 out of 17 children). Out of the six children in the full demonstration condition who built a level-3-construction, two children (aged 4 years 7 months and 4 years 11 months) produced a very similar tripod to the one they saw demonstrated. A further two children in the full demonstration condition built modified tripods with more than three legs. Three additional children in this condition built smaller tripods with a stick level height of two sticks. The four level-3-tripods represented the tallest constructions across conditions (40.5, 42, 43, and 44 cm; M = 42.37 cm, SD = 1.49 cm; compared to the rest of the full demonstration condition towers with M = 18.54 cm, SD = 9.40 cm, and to the baseline towers with M = 15.56 cm, SD = 8.18 cm). Pictures of all towers are presented in Appendix 3.

Table 7. Height and shape of children's towers with the maximum height.

Tower height in levels	Tower shape	Shape description	Condition: Baseline	Condition: Full demo
Level 3	Tripod	Three legs, combined with plasticine, two sticks on top of each other added above		2
	Modified tripod	Tripod with more than three legs		2
	Level-3-tower	Three sticks combined vertically on top of each other	1	1
	Other level-3-construction			1
Level 3 total			1/17 (5.9%)	6/17 (35.3%)
Level 2	Level-2-tripod	small tripod (three legs – plasticine – stick)		3
	Level-2-tower	Two sticks combined vertically on top of each other (at least one stick per level)	1	6
	Other level-2-construction		4	
Level 2 total			5/17 (29.4%)	9/17 (52.9%)
Level 1	Level-1-tower	Ball of plasticine with vertical stick on top or two level-1-towers combined with sticks combined at top	1	2
	Hedgehog	Ball of plasticine from which several sticks protrude upward and/or sideward	4	
	Other level-1-construction		2	
Level 1 total			7/17 (41.2%)	2/17 (11.8%)
Level 0	Crashed construction	Any vertical construction (e.g. level-2-tower) which failed to stand alone	2	
	Horizontal construction	Construction with sticks and plasticine, intentionally built in horizontal fashion	1	
	Plasticine tower	Plasticine-only tower	1	
Level 0 total			4/17 (23.5%)	0/17 (0%)

4.2.3.3 Similarity to tripod

The LMM (see above) showed that condition had only a marginally significant effect on the rated similarity of children's towers to the demonstrated tripod, $\chi^2 = 3.049$, df = 1, $p = .081$. However, note that this analysis also included the final towers for children whose tower with the maximum height was built earlier in the trial and did not survive until its end. This was the case for 13 children, 9 of which were in the full demonstration condition. When we excluded these (smaller) towers and reran the analysis, we did find a significant effect of condition on the rated similarity of children's towers to the demonstrated tripod, $\chi^2 = 4.545$, df = 1, $p = .033$, with children in the demonstration condition building constructions that were more similar to the tripod than constructions made by baseline children.

4.2.4 Discussion

This study established that the tower task is suitable for studying the transmission of cumulative technological design in children aged 4 to 6 years. Upon observing a demonstrator build a tripod, i.e., an artefact that children this age would not be able to make on their own, children produced towers that were on average taller than those built by a baseline group who did not have the opportunity for social learning. In addition, some of the children in the demonstration condition produced a faithful copy of the tripod. These findings thus provide the first evidence that 4-year-old children are able to copy cumulative technological design.

However, it needs to be noted that we do not know whether children in the demo condition would also have built taller towers than the baseline if they had been shown a small tower or an inefficient design. If we added such a condition and if we found that children performed better than baseline even when they saw an average or worse-than-average tower, then we would need to conclude that the effect of children building taller towers is due to a *generalized* demonstration effect that results from showing children *any* example of a tower and that has nothing to do with the specific tripod shape. Yet, the results of the study presented in chapter 5 show that when children are presented with average or small towers made by other children, they are not able to improve their tower height beyond what children in a baseline condition achieve. These results suggest that seeing inefficient designs or small towers does not suffice for children to build taller towers than baseline – children would need to be presented with a tall tower and/or efficient design rather than any tower.

Since this study was our first attempt and – to our knowledge – the first study in general to investigate the transmission of cumulative technological design in children, we aimed to replicate our findings in study 2. Another reason for carrying out a replication was the fact that the data for the two conditions in the current study were not collected simultaneously and that they were collected at different places (baseline first in nurseries, full demonstration condition second at Science Museum). This might have made the two groups of children less comparable. A third reason for carrying out study 2 was that we aimed to include several improvements to the study design: We chose a more interactive warm-up game (spinning tops), we removed the turn-taking aspect in the demonstration condition, we changed the wording in the instructions to “tall” and “taller than”, we included a control question to ensure children’s understanding of “taller than”, we placed a box behind the construction table where we put the demonstration tripod (rather than placing it on the floor), and we introduced a simple method that allowed us to measure tower height throughout the trial without interrupting children's building process. Another reason for carrying out study 2 was that we aimed to include an *endstate demonstration condition* to investigate what kind of information children require to benefit from the cumulative technological design demonstration.

4.3 Cumulative culture study 2

4.3.1 Introduction

Study 2 had three goals: First, to replicate the findings from study 1: i) that children in a baseline condition would not produce a tripod on their own, and ii) that children in a full demonstration condition would be able to copy the tripod design. This replication was important because in Study 1 we did not collect the data for the two conditions simultaneously and thus comparability of these two conditions might have been impaired. In addition, and more generally, the reproduction of novel findings is crucial to enhance their credibility – especially when a study is the first of its kind (Open Science Collaboration, 2015). Second, we aimed to investigate whether children would also be able to copy cumulative technological design when they lacked information about the actions involved in producing the design. For this, we added to the baseline and full demonstration condition a third condition, an endstate demonstration. Third, we introduced small improvements to our methodology.

4.3.2 Methods

4.3.2.1 Participants

Seventy-three children (34 boys) between 4 years 2 months and 5 years 8 months (M_{age} = 5 years 0 months, *SD* = 4.31 months) were tested in nurseries and a science museum in a metropolitan area in the UK. The ethnic composition was 59% Caucasian, 27% Asian, and 14% Black. Written informed consent was obtained by participants' parents or guardians prior to the study. Children were randomly assigned to the baseline (*n*

= 23, 43.5% male), full demonstration (n = 23, 56.5% male) or endstate demonstration condition (n = 27, 40.7% male); comparable numbers of children from each testing site were represented in each condition. There were no differences in the distribution of age (Kruskal-Wallis-test, $\chi^2(2) = 0.963$, $p = .618$) between conditions. Another two children were excluded from the analysis as they did not answer the control question correctly.

4.3.2.2 Material

For the warm-up, E placed five plastic spinning tops (diameter 3.5 cm) on the table and invited children to play together. To investigate whether participants understood the meaning of "taller", we included a control question after the warm-up game. E presented children with two Playmobil giraffes of differing sizes (adult giraffe and calf, height 15.5 and 8 cm) and asked them to show her the taller animal. For the construction task, the materials were the same as in study 1, with the following additions: In the demonstration conditions, a box with the same height as the table at which children were building was placed around 20 cm behind the table. This was where E moved the tripod to after presenting it to the children on the table (rather than placing it on the floor). In addition, in order to measure tower height throughout the trial, we used an expandable folding rule attached to the table, opposite of where E was sitting. The measurement was done by visual judgement, a procedure shown to be sufficiently reliable (see Appendix 4). Tower height at the end of the trial was measured with a loose folding rule held right next to the construction.

4.3.2.3 Procedure

The procedure was largely the same as in study 1. After the warm-up, we tested children's understanding of the concept "taller" with a control question for which children were shown two Playmobil giraffes of differing sizes and asked to indicate which animal 0was taller. Children who did not answer correctly also proceeded to the construction task but their data were excluded from the analysis. For the construction task, E said: "The game is to build something that is very tall, as tall as you can make it". She presented children with the material as she did in study 1.

In the demonstration conditions, E said: "Before you start, let me show you what I did earlier!" In the full demonstration condition E built the tripod (~50-60 sec). Upon completion, she said "Finished!" and looked at the tripod for 5 sec. She then placed the tripod on the box behind the table, where it was available for inspection throughout the trial. In the endstate demonstration condition, E fetched a board with a ready-made tripod from behind a barrier standing next to her and placed it on the table. She looked at it for 5 sec and moved it to the box. The rest of the instructions in the demonstration conditions was the same as the instruction children in the baseline were given: E said: "You don't have much time to build something that is as tall as possible"; this was to induce them to be quick as their building time was only 6 min. E then encouraged children to start building.

During the building phase, E took measurements of children's towers using a folding rule attached to the table, each time participants made an addition to their construction which increased its height and if the construction was standing on its own (i.e., children did not stabilise it with their hands). When the time was up, children were not allowed to touch the construction anymore. Children who held their construction in their hands were asked to place it on the table and those who stabilised it with their hands were told to let go. Towers that could not stand on their own had to be placed horizontally on the table. Tower height was measured again with a loose folding rule held right next to the construction. Once each participant had left the room, E took pictures of the construction (one from each side).

4.3.2.4 Coding and analysis

We were interested in whether children in the demonstration conditions would be able to copy the demonstration tripod. In addition, we investigated whether children in the demonstration conditions would build taller towers than children in the baseline, and how similar children's constructions were to the demonstrated tripod. Again, we measured three variables: *tower height* (height of the tallest construction a participant built), *tower shape* of the tower with the maximum height, and *similarity to tripod* (similarity of the construction to the tripod).

Tower height

Tower height was measured several times: throughout and at the end of the trial. This allowed us to identify each participant's tallest construction, even if the construction did not "survive" until the trial end, e.g., because children disassembled it or because it collapsed due to being too instable or because children tried to further modify their construction. Since instances of tower collapses often resulted from the fact that we encouraged children to use the full building time (i.e., even when children announced they were finished we encouraged them to continue building in order to ensure equal construction time among participants), we measured tower height continuously to ensure that we made a fair evaluation of children's performance. Consequently, for some children tower height represented the height of the tower which stood on the table at the end of the trial, whereas for other children tower height represented the height of a tower that they had built during the trial, but that did not survive until the end.

We analyzed whether maximum tower height differed between conditions, using a multiple regression including condition (baseline, full demonstration, endstate demonstration) as IV, sex (dummy-coded), and age (covariate) as control variables, but no interaction as we did not predict one. Prior to fitting the model, we confirmed that tower height and age had symmetrical distributions. We z-transformed age to a mean of zero and a standard deviation of 1 in order to facilitate the interpretation of the coefficients. We checked the following model diagnostics: normal distribution and residuals plotted against fitted values (to check for homoscedasticity of residuals), DFFits and DFBetas, Leverage, Cook's distance, Generalized VIF, and Levène's test of equal error variances. There were no obvious deviations from the model assumptions. To determine the effect of condition, we compared the fit of the full model with the fit of a model lacking condition as a predictor. The model and the diagnostics were run in R (version 3.2.3, R Core Team, 2013), the Generalized VIF was calculated with the function "vif" of the R package "car" (Fox & Weisberg, 2011). Sample size for the analysis was 73; the alpha level for all analyses was .05.

Tower shape

The shape of children's tallest tower was coded offline based on photos and video stills in the same manner as in study 1.

Similarity to Tripod

Similarity to tripod was coded in the same way as in study 1. The ratings of the two coders correlated significantly ($r = .828$, $p < .001$) and the strength of the relationship between the two ratings was similar to the one in Caldwell & Millen (2008a).

To determine whether the towers in the demonstration conditions were rated as more similar to the tripod than the towers in the baseline, we fitted a LMM (Baayen, 2008) using the function lmer of the R-package lme4 (D. Bates et al., 2013). We used an LMM rather than a Kruskal-Wallis-test (non-parametric version of ANOVA) because we aimed to account for the fact that data points were not independent: For some participants, two towers were entered (the one with the maximum height and the final tower) and raters made 89 judgements each. Into the LMM, we included "rating" as the DV, condition as a fixed effect and random intercepts for participant and rater. We also included a random slope for condition (manually dummy-coded) on rater (D. J. Barr et al., 2013). To allow for a likelihood test, we fitted the model using Maximum Likelihood (rather than Restricted Maximum Likelihood, Bolker et al., 2008). We checked for normal distribution and homoscedasticity of the residuals by visually inspecting a qq-plot and the residuals plotted against the fitted values and found no obvious violation of these assumptions. The significance of condition was determined by comparing the full model against a reduced model (lacking the variable condition) by a likelihood ratio test (R function "anova" with argument test set to "Chisq", Forstmeier & Schielzeth, 2011). Post-hoc Tukey tests were carried out using the R package multcomp (Hothorn et al., 2008). The sample size for this model was 178 ratings made on 89 towers (two ratings per tower).

4.3.3 Results

4.3.3.1 Tower height

Across conditions, children's average tower height was 25.05 cm (*SD* = 11.68 cm), ranging from 0.3 to 45.5 cm. In the baseline, tower height was on average 17.84 cm (*SD* = 10.60 cm, range 0.3-41.0 cm); in the full demonstration condition, average tower height was 28.07 cm *(SD* = 10.96 cm, range 15-45.0 cm); in the endstate demonstration condition it was 28.62 cm *(SD* = 10.68 cm, range 2-45.5 cm). Pictures of the towers are in Appendix 5.

Overall, sex, age in months, and condition together had a clear effect on tower height (comparison full-null model: $F(4, 68) = 5.339$, $p < .001$). Specifically, tower height was significantly affected by condition ($F(2, 68) = 7.992$, $p < .001$; $R^2 = .19$): Compared to children in the baseline, children in the full demonstration condition built towers that were on average 10.65 cm (*SE* = 3.12; 95% CI [4.43; 16.87]) taller ($p = .001$), and children in the endstate demonstration condition built towers that were 10.51 cm (*SE* = 2.98; 95% CI [4.56; 16.45], $p < .001$) taller than towers in the baseline. Tower height did not differ between the two demonstration conditions (estimate+SE: -0.14+3.02, $t_{68} = -0.047$, $p = .962$). Note that even though at the group level, children in the demonstration conditions built taller towers than children in the baseline, only 10 out of 52 children in the demonstration conditions were able to build towers that went *beyond* the maximum height achieved in the baseline.

Age also had a significant, but small positive effect on tower height: With each standard deviation increase in age (4.31 months), average tower height increased by 2.91 cm (*SE* = 1.25, $t_{68} = 2.33$, $p = .022$; $R^2 = .07$; Fig. 9). Sex had no effect on tower height

(0.30+2.50, t_{68} = 0.12, p = .906). Thus, results showed that older children built on average taller towers than younger children and that children in both demonstration conditions built taller towers than those in the baseline.

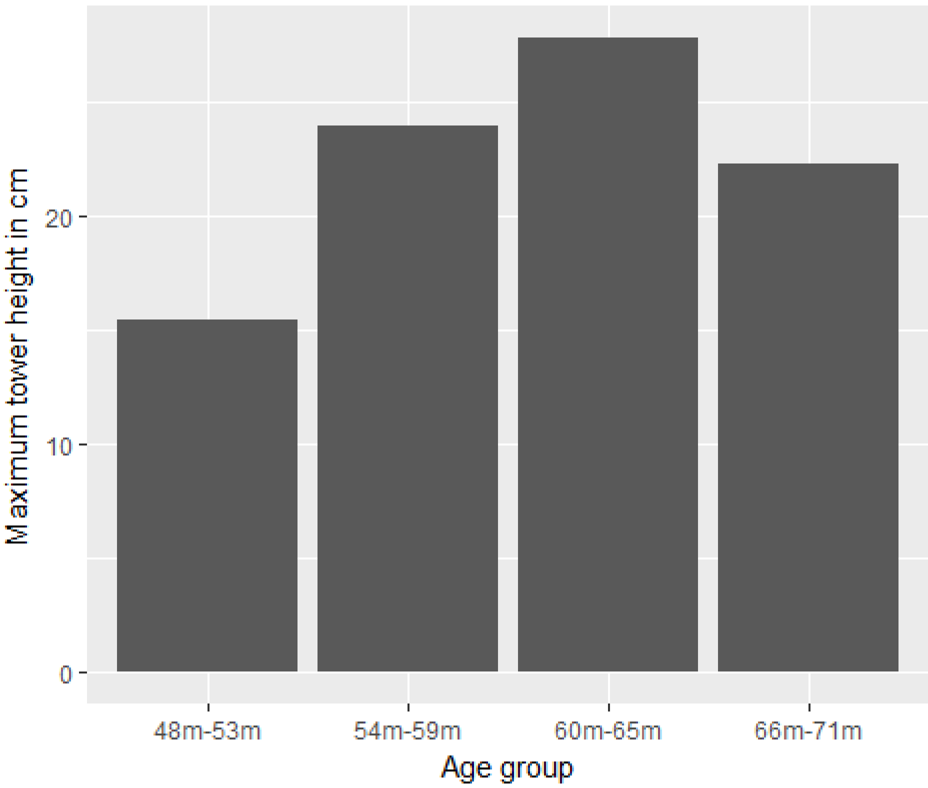


Fig. 10. Maximum tower height in cm displayed by age groups (4-4.5 years, 4.5-5 years, 5-5.5 years, 5.5-6 years).

4.3.3.2 Tower shape

No child in the baseline made a tripod, thus replicating the finding from study 1. The most common tower shape was a level-1-tower in the baseline and a level-2-tower in both demonstration conditions (Table 8). Crucially, children in both demonstration conditions built tripods: Three children in the full demonstration and one child in the endstate demonstration condition copied the tripod (three of these children were younger than 5 years). A further four children built smaller versions of tripods (level-2-tripod). Interestingly, one child in the endstate demonstration condition built a level-4-tripod. Although this tower exceeded the demonstration tripod with regard to height in stick levels, the height in cm did not exceed the tripod height as the child's tower was somewhat crooked.

4.2.3.3 Similarity to tripod

Mean similarity of baseline towers to the tripod was 2.12 (*SD* = 1.08); for full demonstration towers it was 3.05 (*SD* = 1.44), and for endstate demonstration towers 3.05 (*SD* = 1.61). Condition had a significant effect on the rated similarity of children's towers to the demonstrated tripod, $\chi^2 = 7.934$, df = 2, $p = .019$. Post-hoc Tukey tests showed that towers in the full demonstration condition were rated significantly more similar to the tripod than baseline ($p = .036$) and the same pattern was found for the towers in the endstate demonstration condition ($p = .023$). Ratings for the towers in the two demonstration conditions did not differ ($p > .999$).

Table 8. *Distribution of tower height (in stick levels) and shape in the three conditions.*

Tower height in stick levels	Tower shape	Shape description	Condition Baseline	Full demo	Endstate demo
Level 4	Level-4-tripod	As Tripod, but additional stick on top			1
Level 4 Total			0/23 (0%)	0/23 (0%)	1/27 (3.7%)
Level 3	Tripod	Three legs, combined with plasticine; two sticks on top of each other added above		3	1
	Level-3-tower	Three sticks combined vertically on top of each other (at least one stick per level)	2	3	6
Level 3 Total			2/23 (8.7%)	6/23 (26.1%)	7/27 (25.9%)
Level 2	(modified) Level-2-tripod	small tripod (at least three legs – plasticine – stick)		2	2
	Level-2-tower	Two sticks combined vertically on top of each other (at least one stick per level)	4	6	8
	Other level-2-constructions		1	2	3
Level 2 Total			5/23 (21.7%)	10/23 (43.5%)	13/27 (48.1%)
Level 1	Level-1-tower	Ball of plasticine with vertical stick on top or two level-1-towers combined with plasticine at top	6	5	4
	Hedgehog	Ball of plasticine from which several sticks protrude upward and/or sideward	5	2	
	Other level-1-construction				1
Level 1 Total			11/23 (47.8%)	7/23 (30.4%)	5/27 (18.5%)
Level 0	Horizontal construction	Construction with sticks and plasticine, intentionally built in horizontal fashion	3		
	Plasticine tower	Plasticine-only tower	2		1
Level 0 Total			5/23 (21.7%)	0/23 (0%)	1/27 (3.7%)

4.3.4 Discussion

Study 2 investigated 1) whether children would be able to use social and/or technical information provided by a demonstration of a cumulative technological design (tripod) in order to improve their tower building skills and 2) whether some of the children would also copy the cumulative technological design even without having access to action information. 4-6-year-olds were assigned to either a baseline, a full demonstration or an endstate demonstration condition and were asked to build something as tall as possible out of plasticine and sticks. Children in the full demonstration condition observed E build a cumulative technological design (tripod) and children in the endstate demonstration condition were presented with a ready-made tripod.

We replicated study 1 by showing that children in the baseline did not invent the tripod shape; in addition, they neither reached the demonstrated tripod height nor went beyond it. We also found that children in both demonstration conditions built on average taller towers than children in the baseline, suggesting that children were able to pick up task-relevant information from the cumulative technological design demonstration and to improve their performance. This is in line with a study on adult social learning showing that access to social information can boost a learner's performance (Derex & Boyd, 2015).Yet, even though towers in the demonstration conditions were on average taller than baseline towers, there was some overlap between the conditions. This contrasts with our finding with regard to tower *design*, where there was a clear difference between the baseline and the demonstration conditions (no tripods built in the baseline). Thus, the difference between baseline and demonstration conditions with regard to tower height was less dramatic than the difference with regard to tower design.

Crucially, we found that some children (including 4-year-olds) in both demonstration conditions copied the specific and efficient shape of the tripod, even though children were not instructed to do so. Ratings by independent coders also showed that the towers in the demonstration conditions were more similar to the demonstrated tripod than were the towers in the baseline. Thus, this study demonstrates for the first time that 4- to 6-year-old children are not only able to use social and technical information to improve their performance, but that some of the children also spontaneously copy an efficient cumulative technological design. We do not claim that the ability to copy tripods is present in *all* children of this age (after all, individuals' ZPDs differ; yet it may be the case – given that we did not tell children to copy the tripod, but just asked them to make something as tall as possible). As children did not receive an instruction or incentive to copy the tripod, our study was a conservative test of whether children can copy cumulative technological design and so it was even more impressive to find that some children indeed copied the tripod.

Intriguingly, children in the endstate demonstration condition did not perform differently from children in the full demonstration condition: Despite lacking information about the building process of the tripod, children in the endstate demonstration condition still copied the tripod, suggesting that they were able to recreate the tripod via emulation learning: Our participants were able to identify the necessary building steps from looking at the tripod and to recreate it through reverse-engineering. This is evidence that 4- to 6-year-olds do not necessarily require action information to be able to copy cumulative technological design (i.e., technological culture that is beyond their own spontaneous abilities). Rather, children were able to use endstate emulation as a social learning mechanism to copy the cumulative technological design in our task; copying of the actions involved proved not necessary.

One might ask whether our finding suggests that mere stimulus enhancement might also be sufficient to enable the transmission of culture-dependent traits. In our study, stimulus enhancement would be defined as a change in children's behaviour after the actions or products of a demonstrator have attracted children's attention towards a stimulus (the tripod); no attention to the demonstrator's goal is required (Hoppitt & Laland, 2013). In contrast, while endstate emulation also draws the learner's attention towards a stimulus or results of a behaviour, it more specifically describes a situation in which the attention to a stimulus (the tripod) elicits a *specific* change in children's behaviour, namely in that they will be more likely to produce the observed endstate. Learners not only attempt to recreate observed results, but also *goals* (Carpenter & Call, 2002). Given the design and findings of our study (presenting children with both goal and result information and finding at least some children copied the tripod), we think that children's learning in the demonstration conditions is best described as emulation rather than stimulus enhancement. Of course, the demonstration of an endstate not only activates emulation learning, but potentially also stimulus or local enhancement. However, in how far stimulus enhancement on its own would allow for transmission of a culture-dependent trait would have to be investigated specifically via conditions that can disentangle the different effects. For example, one condition could present children with a tripod, but no or a different goal (e.g., building a construction that a plush toy could hide under). Our prediction would be that simply drawing children's attention to the tripod would not suffice for them to copy it. Going away from our specific task to the question whether stimulus enhancement in general could enable the acquisition of culture-dependent traits, we would argue that in most, if not all, cases this would not be possible as learning by stimulus enhancement consists of a considerable amount of individual invention: every element apart from the knowledge on which stimulus to act would have to be added individually. In contrast, culture-dependent traits are by definition those that the individual in question would not be able to invent on its own, making copying (of actions and/or results) a necessary requirement for acquiring culture-dependent traits.

Even though children in our study were not instructed to copy the tripod, it is worth considering that despite receiving action or endstate demonstrations, only a few children actually reproduced the tripod. A number of factors might explain this result. First, some children might have lacked the motivation to copy the tripod. After all, there was no direct incentive for copying (other than it being tall, of course). In addition, the task was not framed within a competitive context, so children might have preferred to not copy explicitly, but to realize their own ideas and/or to explore the materials. For example, we had a couple of children who – upon receiving the demonstration – acknowledged that the tripod was “a good tower”, but then went on making something completely different; in particular, one boy said at this point: “Oh, this is good, but I will make something else”. Second, still developing fine motor abilities at this age probably also contributed to the fact that only few children copied the tripod. Finally, difficulties in causal reasoning and insight into the mechanics of the construction task probably also played a role for children’s low copying rates. In general, children did not yet have a full understanding of the basic physics involved in the task, as several children used only little plasticine for the stand and/or too much for the joints further up. It has been shown that 5-year-olds have difficulty using diagonals when building towers and that this is due to a still developing ability to combine two axes of a coordinate system (Frye, Clark, Watt, & Watkins, 1986; Gentner et al., 2015). Given this difficulty, at least some of our participants might have failed to explicitly recognize the tripod legs as an especially stable tower base. Despite these limiting factors, however, we still found that children were able to pick up useful knowledge and strategies from the demonstrations, resulting in them making taller towers than the baseline children – and some did indeed successfully copy the tripod design.

4.4 General discussion

In this section we will discuss a potential criticism against the studies presented in this chapter: One might argue that the absence of the tripod in the baseline conditions of studies 1 and 2 does not necessarily mean that the tripod represents cumulative technological design for 4- to 6-year-old children (i.e., a technological product that is within the ZPD of children that age). It might be possible that even though the tripod did not occur in the baseline, it is still within children's spontaneous abilities (ZAD), but that environmental and/or motivational reasons prevented children from showing their "true capacities". We think this is unlikely to be the case because 1) children were sufficiently motivated to do the task (they were aware that they could win a sticker; the vast majority of children used the full building time; we encouraged children throughout the task to make their construction "even taller"), 2) they had sufficient material at their disposal, and 3) the finding that children in the baseline did not spontaneously build a tripod was shown in two independent studies (studies 1 and 2). As indicated above, participants had limited knowledge about the physics involved in the task, which makes it further unlikely that 4- to 6-year olds can invent the tripod on their own.

However, one might also argue that the fact that children in the baseline did not build a tripod is due to our specific instructions: We merely asked children to build "something tall", and so we might have increased children's awareness of the height aspect only without also directly addressing the importance of the shape aspect. Given different instructions (e.g., to build "something tall *and sturdy*"), one might expect some baseline children to actually make tripods. However, we would not expect this to happen as in our view the instruction to build something tall does not deemphasize the shape aspect. This is because height and shape are closely connected: When aiming to build a tall construction, one almost automatically needs to take the shape aspect into consideration, for some shapes will be more suitable for building a tall tower than others. Nevertheless, we acknowledge that we currently do not know how baseline children would perform if given different instructions. Future studies are needed to address this question. Note, however, that the studies presented here were not designed to answer questions related to the instructions, but to test children's capacity of making a tripod given an instruction to make something tall in the presence or absence of a tripod demonstration.

Turning to a more general discussion of the studies presented in this chapter, we would like to emphasize that researchers studying the emergence and transmission of culture-dependent traits need to apply clear criteria for determining the presence of those traits. Previous studies (Caldwell & Millen, 2008a; Dean et al., 2012) have not included control conditions, making it "difficult to conclude with certainty that these experiments have demonstrated true cumulative culture" (Mesoudi & Whiten, 2008, p. 3494). While not giving us absolute certainty, including a baseline condition helps determine whether a trait (e.g., building a tripod) can be easily generated by or tends to be beyond the spontaneous capacity of individuals (see e.g. Osiurak et al., 2016; Zwirner & Thornton, 2015). Our studies are the first to include a baseline condition in the tower task. This allowed us 1) to assess children's spontaneous, asocial learning capacities, 2) to identify a tower representing cumulative technological design for 4- to 6-year-olds (tripod), and 3) to make meaningful comparisons of the performance in the demonstration conditions with baseline behaviour. Even though we could not fully rule out that some children had previous experience with a similar game, the weak baseline performance (low height, no tripods) is reassuring that the task was sufficiently novel (and difficult) for our participants. We thus established that the tower task is suitable for studying the transmission of culture-dependent traits in young children: 4- to 6-year-olds are old enough to possess the necessary fine motor abilities to carry out the task, but are still young enough to be able to benefit from social information.

Our results are in line with some adult studies claiming that endstate emulation is capable of transmitting cumulative cultural information (Caldwell & Millen, 2009; Caldwell et al., 2012; Zwirner & Thornton, 2015). However, it should be noted that in the study by Zwirner and Thornton (2015) baseline performance was not significantly different from demonstration conditions; in the studies by Caldwell and colleagues it is likely that participants brought some previous cultural knowledge about how to make towers/paper planes to the task, so that cumulative cultural intelligence might have been a possible confounder in these tasks.

In addition, it has also been argued (Boyd & Richerson, 1996; Tomasello, 1999a) and demonstrated experimentally in adult studies (Derex & Boyd, 2015; Derex et al., 2012; Morgan et al., 2015; Schillinger et al., 2015; Wasielewski, 2014) that seeing only an end product may *not* be sufficient for transmitting cumulative culture. One factor possibly influencing whether culture-dependent traits can be transmitted via emulation is the cognitive transparency of the product in question (Caldwell & Millen, 2009; Derex & Boyd, 2015); however, the mechanisms of this dependency are still unclear. It seems that cultural traits exhibiting sufficient cognitive transparency may be acquired via emulation, because seeing the end product allows learners to reproduce them via reverse-engineering (e.g., building paper planes, Caldwell & Millen, 2009; spaghetti towers, Caldwell et al., 2012; baskets, Zwirner & Thornton, 2015; or plasticine-and-stick tripods (our studies)). Similarly, Want and Harris (2002) proposed that if children have sufficient knowledge about the affordances and properties in a task, its "solution can be emulated" (p. 12), otherwise children would need to revert to copying actions. In contrast, the transmission of cultural products that are cognitively more opaque (e.g., novel weight-carrying devices, Wasielewski, 2014; virtual fishing nets, Derex et al., 2012; foam handaxes, Schillinger et al., 2015; stone tools, Morgan et al., 2008; or complex (virtual) totems, Derex & Boyd, 2015) seems to require additional information about the movements of the objects involved and/or of the bodily actions of the demonstrator, so that learners would be able to copy the trait via object-movement reenactment (a fine-grained version of emulation learning, Custance, Whiten, & Fredman, 1999), action copying (roughly: imitation, Whiten et al., 2009) or both (Acerbi & Tennie, 2016). Therefore, it might be possible that cumulative cultural products can be differentiated by their cognitive transparency, with this transparency in turn influencing whether emulation (learning from results only) is sufficient to transmit the given cultural product. Further studies are needed – both in human children and adults – to identify whether and under which conditions emulation can transmit culture-dependent traits.

Finally, we would like to address two potential points of criticism about the design of our study. First, we need to acknowledge that we do not know whether children's performance in the demonstration conditions – building taller towers and copying the tripod – was due to them being given a "height goal" question or whether we would have found the same result of children were shown the tripod but asked a different question for which the tripod was not the most efficient tower shape, e.g., when asked to make a tower that is best for a plush toy to hide under or that a friend will like best. We do not know whether children would respond differently in these conditions. We would predict them to do indeed behave differently, i.e., to not – or to at least less often – copy the tripod as we would assume that children were sensitive to the specific goal presented and would be able to evaluate the presented tower with regard to this goal. However, if we found that children copied the tripod equally readily in these additional conditions we would tend to conclude that children rather indiscriminately copy a presented culture-dependent product irrespective of the specific goal presented. It could be possible that the basic context that is present in all these conditions – the child being in a one-to-one situation with an adult who is showing them something – is interpreted by the children as a pedagogical context in which they are expected to copy what E showed them (Csibra & Gergely, 2006). This would then be further evidence for children's sensitivity to cultural learning contexts.

Second, we neither know how children would have performed if given a demonstration of a different tower, specifically one that was smaller than average baseline performance. It is currently unclear whether children would still be able to benefit from this demonstration and build taller towers (something we would not expect them to do given the findings of the study presented in chapter 5), which would be evidence of a generalized demonstration effect, or whether children would rather tend to copy this tower and thus build significantly smaller constructions than in the current study. Further research including conditions with different demonstrations is needed to investigate these possibilities.

4.5 Summary and link to chapter 5

This chapter showed that the tower task can be used to study the transmission of cumulative technological design in children. It also provides the first evidence that young children indeed learn from, and that some of them even copy, cumulative technological design, and that – in line with claims from some adult studies – action information is not always necessary to learn from and even copy cumulative technological design.

These findings raise the question whether young children, at least some of whom our studies have shown to be capable of copying cumulative technological design, would also be able to further transmit and maintain cumulative technological design among themselves (i.e., across "generations"). Studies applying the transmission chain paradigm are capable of exploring whether cumulative technological design can be maintained along a chain of children or whether it would disappear (Mesoudi, 2007; compare with the method applied in Morgan et al., 2015). Another question is whether children would also be able to *produce* cumulative technological design by themselves, i.e., whether they can successively build upon ever better solutions, thus exhibiting a ratchet effect. It might be that children find this innovative aspect of the ratchet effect difficult. Support for this thought comes from our finding that participants in our demonstration conditions were not able to go beyond the height of the demonstrated tripod and that only one participant (in the endstate demonstration condition) built a level-4-tripod. Furthermore, research on innovation in children demonstrated surprisingly low innovation rates with regard to making tools (Chappell, Cutting, Apperly, & Beck, 2013) or inventing strategies to retrieve more rewards from a puzzle-box (Carr, Kendal, & Flynn, 2015). These studies suggest that children well into their primary school years seem to struggle with inventing (better) solutions to new tasks. This might imply that groups of young children might not yet be able to show a ratchet effect as continuing limitations on their ability to innovate represents a critical bottleneck for the development of a capacity for producing culture-dependent traits (Carr, Kendal, & Flynn, 2016; Enquist et al., 2008). Again, the transmission chain method is a powerful means for investigating whether generations of children can produce culture-dependent traits by adding innovations to existing solutions and then transmitting these. Chapter 5 presents a study in which we use the transmission chain method to examine whether groups of young children exhibit a ratchet effect in our tower task and are able to produce cumulative technological design.

CHAPTER 5: YOUNG CHILDREN FAIL TO GENERATE A RATCHET EFFECT IN A TOWER CONSTRUCTION TASK

5.1 Introduction

The ability to produce cumulative culture – i.e., cultural traits (artefacts, skills, knowledge) that could not have been created within a single lifetime but that instead are the result of a gradual evolution over time (Boyd & Richerson, 1996) – is argued to rely on high-fidelity social transmission such as imitation and imitation-based teaching, and a capacity for innovation as the main – but perhaps not the only – cognitive factors, alongside fostering social and demographic factors, such as group size and the interconnectivity of social networks (Caldwell & Millen, 2009; Dean et al., 2014; Derex & Boyd, 2015; Derex et al., 2013; Enquist et al., 2008; Enquist, Strimling, Eriksson, Laland, & Sjostrand, 2010; Henrich, 2004; Hill, Barton, & Hurtado, 2009; Hill et al., 2011; Kempe & Mesoudi, 2014; Kline & Boyd, 2010; Lewis & Laland, 2012; Migliano et al., 2017; Muthukrishna & Henrich, 2016; Muthukrishna, Shulman, Vasilescu, & Henrich, 2014; Powell et al., 2009; but see Morin, 2016a). Which social learning mechanisms are faithful enough to create and maintain cumulative culture is still under debate (Caldwell, Atkinson, & Renner, 2016, see chapter 4).

A major advance in this regard has been the introduction of the transmission chain method as an experimental design to simulate cumulative cultural evolution under controlled conditions in the laboratory – because these allow examining and comparing the down-the line effects of different social learning mechanisms (Caldwell & Millen, 2008a, b). In these studies, a series (a "chain") of participants are sequentially asked to take part in a certain task (e.g., building a paper plane or a weight-carrying device; Caldwell & Millen, 2008a; Wasielewski, 2014). Participants in each position of the chain (apart from the first position) are presented with socially generated information, e.g., with the actions and/or final products (endstates) of participants in earlier positions of the chains. By varying the type of information presented (e.g., actions, endstates, verbal instructions) as well as the combination of these types (e.g., actions *plus* endstates) the transmission chain paradigm allows identifying which social information (or combination of different pieces of information, see Acerbi & Tennie, 2016) is necessary for a ratchet effect to emerge.

Using such transmission chain methods, it has been suggested that action-copying (an important component of imitation) is not always necessary for the ratchet effect to occur: Cultural products with a transparent physical structure (e.g., paper planes, simple baskets) could be reverse-engineered and so emulation seems to be sufficient to produce a ratchet effect in these cases (Caldwell & Millen, 2009; Zwirner & Thornton, 2015). However, the relative ease with which adults produced a ratchet effect in these tasks was probably not only because of their transparent structure but also because participants possibly brought to the tasks some previous cultural knowledge about how to make paper planes or a stable basket. In other words, cumulative cultural intelligence is a possible confound in these tasks, making it difficult to determine whether adult participants were able to produce a ratchet effect via endstate copying only. A powerful way to minimize the influence of such confounds are studies on young children which employ novel or unusual tasks, as done in the current study.

Adult participants have also been tested in cognitively more opaque tasks such as making a weight-carrying device or knapping a stone tool. These studies found that emulation was not sufficient to produce or maintain a ratchet effect (Morgan et al., 2015; Wasielewski, 2014; however, the short trial length in Morgan et al., 2015, might have suppressed an effect of emulation learning). In sum, while research on the conditions under which different social learning mechanisms support cumulative cultural evolution is still ongoing, transmission chain studies have already demonstrated that chains of human adults can produce a ratchet effect in the laboratory, and that imitation (copying goals, actions, and results) and possibly also emulation (copying only goals and results) are faithful enough to support this effect (but see Morin, 2016, for critique of method details).

Developmental studies can shed light on the question whether groups of younger humans are also able to produce traits that they cannot invent individually (culture-dependent traits) or whether this ability is restricted to adults and why or why not this may be. Note that we explicitly do not phrase this question as to whether children can produce "cumulative culture" – instead we prefer to use the broader term "culture-dependent trait" (see chapter 1). Culture-dependent traits describe traits that a given individual or group of individuals at a given developmental stage can only acquire through faithful social learning. Therefore, the definition of culture-dependent traits also entails traits that do not represent cumulative culture (see chapter 1, Fig. 3). For example, in chapter 4 we showed that 4- to 6-year-olds are able to make a tripod only after social learning took place. Does this mean the tripod represents cumulative culture? No, as it might be possible that the tripod could be invented by any independent individual at some later point in the lifespan without social learning, thus violating the definition for cumulative culture. According to the theoretical framework presented in chapter 1, the tripod would then be categorized as a ZLS_{far} trait: For children aged 4 to 6, it can only be acquired through social learning, but humans would in principle be able to invent the trait individually at some point during their lifespan.

As Vygotsky (1978) stated, pairs or groups of children are only able to collaboratively create those traits that are within their immediate “developmental level” (p. 45), i.e., within their ZPD. Thus, in our tower task, we would expect groups of children to be able to invent complex forms of towers that are beyond what children at this age can invent individually but that can be learned socially (such as the tripod), but we would not expect them to invent the Eiffel Tower (i.e., cumulative culture proper) because it is too many steps away from their current development level.

Another reason to not use the term cumulative culture with regard to our study on children is to avoid confusion about nomenclature. Cumulative culture is usually assumed to be created by *adult* members of a group who have acquired reasonably large and deep cumulative cultural intelligence. In contrast, children are thought to primarily learn about and absorb their cultures rather than to add to them (Hill, 2010). However, this is not to say that research on the origins of the ratchet effect in children is futile; we need to know when the ability to produce a ratchet effect occurs in development. In addition, studying young children can help to understand the cognitive and motivational requirements for the ratchet effect because children have not yet gained as much cultural knowledge as adults and so the cognitive processes can be studied more (but of course not completely) independently from cumulative cultural intelligence.

Returning to empirical work on the ratchet effect, human children have only recently become a research focus (Dean et al., 2012; Flynn, 2008; McGuigan & Graham, 2010; Tennie, Walter, Gampe, Carpenter, & Tomasello, 2014; see also chapter 4). So far, only a sub-form of the ratchet effect – the *subtractive ratchet effect* (Tennie et al., 2014) – has been found in children. In her transmission chain study, Flynn (2008) showed that groups of 2- and 3-year-olds “weeded out” causally ineffective actions seeded by an experimenter when retrieving a reward from a puzzle box. Using the same apparatus, McGuigan and Graham (2010) showed a similar effect in 5-year-olds – but not 3-year-olds who instead copied both relevant and irrelevant actions. Tennie et al. (2014) seeded chains of 4-year-olds with inefficient ways of carrying rice from one place to another and also found that ineffective solutions were weeded out along the chains. These studies suggest that chains of children undergo an evolution from producing a suboptimal initial action towards a more efficient one. However, the target behaviours in these studies (e.g., opening a puzzle box or choosing the most efficient types of tools to carry as much rice as possible) were still within children’s spontaneous reach (ZAD, Vygotsky, 1978): Children in baseline groups, who were not exposed to the previous generation’s behaviour and outcomes, produced comparably high efficiency levels individually. Thus, these studies show that children are able to make “upgrades to techniques” (Tennie et al., 2009, p. 2413), i.e., modifications that are still fully within the scope of spontaneous individual invention. Such a form of cultural evolution, individually directed towards a “gravitational centre”, may therefore be better described as a “step-wise tradition” (ibid.) as contrasted with producing culture-dependent behaviours. However, it is the study of the *additive type of ratchet effect* (Tennie et al., 2014), defined as the build-up of cultural improvements beyond a level that could be reached individually and leading to “open-ended” outcomes rather than to asymptotic optimizations (Tennie et al., 2017) – which promises insights into the developmental origins of human cumulative culture.

Two studies so far have aimed to investigate the additive ratchet effect in children: Dean et al. (2012) presented groups of 3- to 4-year-olds with a three-stage puzzle box whose stages had to be completed sequentially. The authors found that the majority of the groups had at least two children solving the third stage of the box and concluded that 3- and 4-year-olds were capable of “cumulative cultural learning” (p. 1117). However, it is an open question whether the box adequately operationalized the production of culture-dependent behaviours: Even though the box technically allowed for an accumulation of the number of steps (opening one, two, three stages), it is unclear whether it also captured an increase in complexity or efficiency beyond a level that could be reached individually: A baseline condition controlling for such asocial invention was missing. Thus, it cannot be ruled out that the complete solution of the task was within the capacity of individual children and would thus represent an example of step-wise traditions. Social support and learning from peers may have facilitated the children’s solutions, but were perhaps not necessary for their success[18].

The rice-carrying study by Tennie et al. (2014) was also designed to investigate an additive ratchet effect in children. In contrast to Dean et al. (2012), this study did include a baseline condition. Compared to the baseline, the authors found no evidence for chains of 4-year-olds being able to generate and spread efficient techniques of rice transport that would have gone beyond the individuals’ performance. However, it could be argued that the task was more suitable for inducing a subtractive rather than an additive ratchet effect as the efficient solution of the task – using the containers and tools provided to carry as much rice as possible – proved to be well within children’s reach. In addition, this task as

[18] Indeed, current research in our lab suggests that children can reach the third level of the box individually (Gwilliams, Reindl, & Tennie, 2017).

well as the other studies investigating the subtractive ratchet effect (Flynn, 2008; McGuigan & Graham, 2010) were asymptotic, i.e., any improvements would eventually converge to a single optimal behaviour (using a certain set of materials to transport the rice or getting rid of all irrelevant actions on the puzzle-boxes). This prevented the occurrence of an open-ended, additive ratchet effect. In sum, there is so far no clear evidence – nor a clear test – for young children being able to produce culture-dependent traits.

In chapter 4 we investigated whether one crucial requirement for the ratchet effect – namely, the ability to copy culture-dependent traits – is already present in young children. Results showed that 4- to 6-year-olds are already able to copy cumulative technological design (the tripod) and that they can do so via emulation. However, whether children would also be able to *produce* (rather than just copy) cumulative technological design is far from certain, especially because several studies indicate that children's innovative abilities seem to be a critical developmental bottleneck (Beck, Chappell, Apperly, & Cutting, 2012; Carr et al., 2015; Chapell et al., 2013; Cutting et al., 2011; Nielsen et al., 2014; Sheridan et al., 2016; Whalley, Cutting, & Beck, 2017).

This chapter investigates – using the transmission chain method – whether children between 4.5 and 6 years are able to produce cumulative technological design in the tower task. We randomly assigned children to one of two conditions: a baseline, used to replicate our results from chapter 4 and with which to contrast children's performance in the second condition, the transmission chain condition. For the transmission chain condition, we implemented eight chains with a length of ten children each. Chains were separated by sex (4 boy chains, 4 girl chains) because a previous diffusion chain study found a gender effect in copying (Flynn & Whiten, 2008). Since the tower task is a rather transparent task for both adults (Caldwell et al., 2012) and children (chapter 4), requiring no action information for participants to copy, the children in the transmission chain condition of the current study received an endstate demonstration: Children were presented with the final tower made by previous children but did not see the building of these towers. Thus, children in both the baseline and in the transmission chain condition were tested asocially, i.e., they had no contact to other children during the entire study. This also rendered the implementation of a transmission chain study with children more practical.

5.2 Method

5.2.1 Task design

Participants were allocated to one of two conditions: the baseline (*n* = 21) or the transmission chain condition (*n* = 80). In the baseline, children were asked to build something as tall as possible using plasticine and sticks, but were not provided with demonstrations, thus replicating the baseline conditions of chapter 4. In the transmission chain condition, children were assigned – in succession – to one position (1 to 10) within a same-sex chain. As in the baseline, children were tested individually and asked to build something as tall as possible; in addition, they had the opportunity for social learning as they were shown the constructions made by their two immediate predecessors (by design, this only happened for children in positions 3 to 10; children in position 2 only saw the construction of the first child in the chain, and children in position 1 did not have access to any social information, thus experiencing the same situation as baseline children).

The presented tower(s) were mainly the original towers that had been built by previous participants tested earlier that day or the day before. However, since testing took place over several sessions spread out over several months, sometimes the end of a session did not coincide with the end of a chain, and so we had to disassemble children's towers and reassemble them at the beginning of a new session. Out of 72 towers that were demonstrated to children (8 chains x 9 towers), E had to rebuild 18 towers from photos of the original towers.

5.2.2 Participants

The final sample consisted of 101 children (51 boys) between 4 years 5 months and 5 years 8 months (M_{age} = 57.87 months, *SD* = 3.40 months) tested in Birmingham, UK (in a science museum (*n* = 56) and a zoo (*n* = 32)) and in a rural area in Germany (in a school building outside of school hours (*n* = 9) and a nursery (*n* = 6)). Participants were recruited via information letters sent to schools and flyers handed out to parents at various events. Written informed consent was obtained by children's caregivers before testing started. Parents were present during testing for all children except for five children tested in the nursery. The ethnic composition of the sample was 62.7% White British, 13.7% White German, 12.7% Asian British, 8.8% other Mixed, 1% other White. Children were rewarded with stickers, a certificate, and a voucher (free entry to a museum or an indoor play area) regardless of success. An additional 16 children were tested but had to be excluded from the study because they did not meet the required age range (*n* = 4, answered the control question wrong (*n* = 3, or due to experimenter error (*n* = 2), and seven children had to be excluded because they were tested in a transmission chain following a child who did not meet the required age range and thus had to be excluded), and seven children had to be excluded because they were tested in a transmission chain following a child who did not meet the required age range and thus had to be excluded.

5.2.3 Material

As in study 2 of chapter 4, we used a set of spinning tops as a warm-up game and two Playmobil giraffes of differing sizes for the control question testing for children's understanding of the concept "taller". For the construction task, we used the same material as in the previous studies, with the exception that children were provided with 100g instead of 70g of plasticine in order to not impose any arbitrary external limits to a potential accumulation in tower design. We used the timer function on an iPad to show children their time progress during the task (analog and digital display). In the transmission chain condition, we also used a movable table (70 x 42 x 18 cm) to present the constructions built by the preceding participant(s).

5.2.4 Procedure

Children were tested individually by the same female E. Both were sitting on the floor, at a low table (h = 15 cm). Opposite of the child, around 2 m away, there was an adult-sized table underneath which there was a moveable table of the same height as the low table the child was sitting at and which E placed the demonstration tower(s) on. A black tablecloth hanging from the big table prevented the view of the constructions at the start of the trial.

At the beginning of the construction task, children in both conditions were presented with the building material (100g of plasticine and 30 plastic sticks) and the goal of the task: to build something "as tall as possible". The demonstration of the material was carried out as in chapter 4. In the transmission chain condition, children were then presented with social information (apart from children in position 1). E said: “So before you start, let me show you what another child/other children did earlier!” E lifted the tablecloth and moved the table presenting the construction(s) closer, so that it was ~ 100-120 cm away from the child. E looked at the constructions for 5 sec, then said: “Now you try to build something that is very tall!” Children in both conditions had 10 min to complete the tasks. In our previous studies we experienced that some children were concerned that they would run out of time, which might have prevented them from concentrating fully on the task. Therefore, in this study we displayed a timer visible to children throughout the task which allowed them to easily track their progress. E also explained that there was no need to rush as the game was not a race. Children were then encouraged to start building. Testing stopped when 10 min were over or when children refused to further engage in the task even after having been encouraged to continue for several times. Most children used the full 10 min (67 out of 101 children); of the 34 children who did not use the full trial length, 28 children were building for 9 min or longer, one was building between 8 and 9 min, two between 7 and 8 min, another two between 6 and 7 min, and one for 5 min 40 sec. In the transmission chains, the towers of the one or two previously tested children were present during the entire building time.

5.2.5 Coding and analysis

For each child, we measured three variables that were live-coded: *tower height*: the height of the tower that was standing on the table at the end of the trial; *tower shape*: the shape of the tower, using the classification system used in chapter 4; and whether or not children produced a *tripod (yes/no)* which is an efficient solution to this task and likely represents a culture-dependent trait for children between 4 and 6 years (see chapter 4).

Unless indicated otherwise, statistical analyses were conducted with IBM SPSS Statistics 24.0. We first analyzed performance of the children in the baseline, describing the heights and shapes of their towers. Then we analyzed tower heights and shapes in the transmission chain condition. We expected to find no difference between the average tower height in the baseline and in position 1 across the chains because children experienced the same test situation (no social information present). We tested this with an independent samples t-test.

5.2.5.1 Research question 1: Was there an increase in tower height across generations?

To investigate whether there was a significant improvement in tower height along the chains – i.e., in order to examine the presence of a ratchet effect – we conducted three analyses: We examined whether the average tower height in position 10 across the chains was significantly different from the average height of the towers in position 1 using a paired samples t-test. Second, in line with Caldwell and Millen (2009), we did the same analysis comparing the average height across the last three towers (position 8, 9, 10) with the average height across the first three towers (position 1, 2, 3). Third, we used a new implementation of Page's L test (Page, 1963; see Stadler, 2015, for evidence that the original test is unsuitable as it is too sensitive) using the R package cultevo (Stadler, 2017) to examine whether there was a monotonic increase in tower height along the chains.

5.2.5.2 Research question 2: Did children produce a culture-dependent technological product?

The analyses for question 1 would tell us whether there was an increase in tower height along the chains. We also aimed to investigate whether this potential increase in performance would lead to the production of culture-dependent traits, i.e., whether children in the transmission chain condition were able to reach tower heights and shapes which baseline children would not reach independently. For this, performance in the transmission chains needs to be contrasted with baseline performance. We compared average tower height in position 10 across all transmission chains with the average height in the baseline using an independent samples t-test. We also investigated whether children in the transmission chains were ever able to build a tripod, i.e., a construction that we assume to be a culture-dependent product for children between 4 and 6 years.

5.2.5.3 Research question 3: Was there cultural variation between the transmission chains?

We were also interested in whether the chains differed significantly from each other in tower shape, i.e., whether we found the emergence of different "cultural lineages" (Caldwell & Millen, 2008a). For this, we first needed comparisons of each transmission chain tower against the remaining nine towers from the same chain (*within-chain similarity*) as well as against all the towers from the remaining seven chains (*between-chain similarity*), resulting in 79 ratings per tower. To get these ratings, we recruited a group of 410 hypotheses-naïve raters on the crowdsourcing platform Crowdflower. The raters were asked to judge the similarity of each tower (both in baseline and the transmission chain, n = 101) against all other towers in our dataset using a 7-point scale from *0* (*not similar at all*) to *7* (*very similar*; points in between were not labelled), totaling in 101 * 100 = 10100 ratings[19]. This procedure resulted in two independent ratings for each pair of images, which allowed us to calculate the reliability of these judgments (see Caldwell & Millen, 2008a). We found a significant correlation between both judgments, $r = .385$, $n = 5050$, $p < .001$. The correlation was low, yet comparable to the correlation found in Caldwell and Millen's (2008a) paper plane task ($r = .387$). As Caldwell and Millen argue, the low correlation coefficient indicates that it was "relatively difficult to objectively judge the similarity of these photographs" (p. 167), especially because such a fine-grained scale was used.

We averaged across the two ratings to calculate the mean similarity of each tower in the transmission chain condition to both 1) the towers within the same chain (n = 9) and 2) the towers of the remaining chains ($n = 70$). A related-samples t-test was used to investigate whether towers were more similar to towers within the same chain than to towers from different chains.

5.3 Results

5.3.1 Baseline performance

Since there were no differences in tower height between boys and girls in either condition, data were collapsed across sex. In the baseline (n = 21), children's constructions had a mean height of 17.20 cm (*SD* = 9.46 cm), ranging from 3 to 32 cm (median = 17.50

[19] Note that even though raters compared all towers in the dataset against each other, we did not use the data comparing baseline and transmission chain towers for any analysis.

cm, mode = 3 cm (n = 3)). Children built level-0-, level-1-, and level-2-constructions, with level-1-constructions being the most commonly reached stick height (42.80%, Table 9). The most common tower shapes were *plasticine-only towers* (level-0-construction, n = 5) and *hedgehogs* (level-1-construction, n = 5). No child built a tripod, thus replicating the finding from chapter 4 that the tripod is likely beyond the inventive power of individual children aged 4 to 6.

Table 9. *Shapes and heights in "stick levels" of the constructions made in the baseline and the transmission chains.*

Tower height in sticks	Tower shape	Shape description	Number of constructions: Baseline	Number of constructions: Transmission chains
Level 4	Level-4-2-leg-tower	Level-4-tower with 2 legs	-	1
	Level-4-tower	4 sticks combined vertically on top of each other (at least 1 stick per level)	-	1
Level 4 Total			-	2/80 (2.5%)
Level 3	Elevated level-3-tower	As level-3-tower, but with increased plasticine base	-	1
	Level-3-tower	3 sticks combined vertically on top of each other (at least 1 stick per level)	-	3
Level 3 Total			-	4/80 (5%)
Level 2	Modified Level-2-tripod	4 or more legs at the base, combined with piece of plasticine at the top, on top of this 1 or more vertical sticks	-	2
	Level-2-3-leg-tower	Level-2-tower with 3 legs arranged in a line	1	-
	Level-2-2-leg-tower	Level-2-tower with 2 legs	-	6
	Level-2-tower	2 sticks combined vertically on top of each other (at least 1 stick	4	15

Tower height in sticks	Tower shape	Shape description	Number of constructions	
			Baseline	Transmission chains
		per level)		
	Other level-2-constructions		1	3
Level 2 Total			6/21 (28.6%)	26/80 (32.5%)
Level 1	Broken level-3-tower	Level-3-tower that broke down to level-1-height	-	1
	Level-1-tower with 2 legs	2 sticks with plasticine base combined with plasticine at the top	-	1
	Level-1-tower with plasticine cap	As level-1-tower, but with pieces of plasticine on top of 1 or more sticks	-	1
	Elevated level-1-tower	As level-1-tower, but with increased plasticine base	2	-
	Level-1-tower	At least 1 ball of plasticine with at least 1 vertical stick on top	2	14
	Hedgehog	Ball of plasticine from which several sticks protrude upward and/or to side	5	21
Level 1 Total			9/21 (42.8%)	38/80 (47.5%)
Level 0	Level-0-tower	Construction involving sticks and plasticine, smaller than height of 1 stick	-	3
	Plasticine-only tower	Construction made from plasticine only	5	2
	Fallen tower	Tower of any height that didn't stand on its own	1	5
Level 0 Total			6/21 (28.6%)	10/80 (12.5%)

5.3.2 Performance in the transmission chains

There were 80 towers (8 chains with 10 towers) with a mean height of 20.86 cm (*SD* = 11.19 cm), ranging from 1.5 to 59.5 cm (mode (*n* = 14) and median were 17.00 cm). As expected, average tower height in position 1 of the chains (*M* = 23.12 cm, *SD* = 6.67 cm) did not differ from the average height reached in the baseline (17.20 cm; independent samples t-test, two-tailed, $t(27) = 1.617$, $p = .118$, Cohen's $d = 0.724$). Children in the transmission chain condition built level-0-, level-1-, and level-2-constructions (as in the baseline), but also level-3- and level-4-constructions (unlike in the baseline; Table 9). As in the baseline, level-1-constructions were most commonly built (47.5%), with hedgehogs being the most commonly made shape (*n* = 21). Whereas the maximum height in the baseline was 32 cm (reached by one child), 7 (8.75 % of the) children in the transmission chains built towers that were taller than this; these children all built level-3- and level-4-towers, apart from one child who built a level-2-tower.

5.3.2.1 Research question 1: Was there an increase in tower height across generations?

We compared average height in position 1 (*M* = 23.12 cm, *SD* = 6.67 cm) to the average height in position 10 (*M* = 17.97 cm, *SD* = 18.43) across all chains and found no difference in height (paired samples t-test, two-tailed: $t(7) = 0.667$, $p = .526$, Cohen's $d = 0.236$). Similarly, there was no difference when average tower height across the first three positions (*M* = 22.04 cm, *SD* = 6.58 cm) was compared with the average tower height across the last three positions (*M* = 19.46 cm, *SD* = 6.96), paired samples t-test, two-tailed: $t(7) = 0.645$, $p = .539$, Cohen's $d = 0.228$). A new, improved implementation of Page's L test (Page, 1963; Stadler, 2015, 2017), testing for the presence of a monotonic increase in tower heights across chains was not significant (*L* = 2309.5, *k* = 8, *n* = 10, *p* = .922), suggesting that average tower height along the chains did not increase (see also Fig. 10).

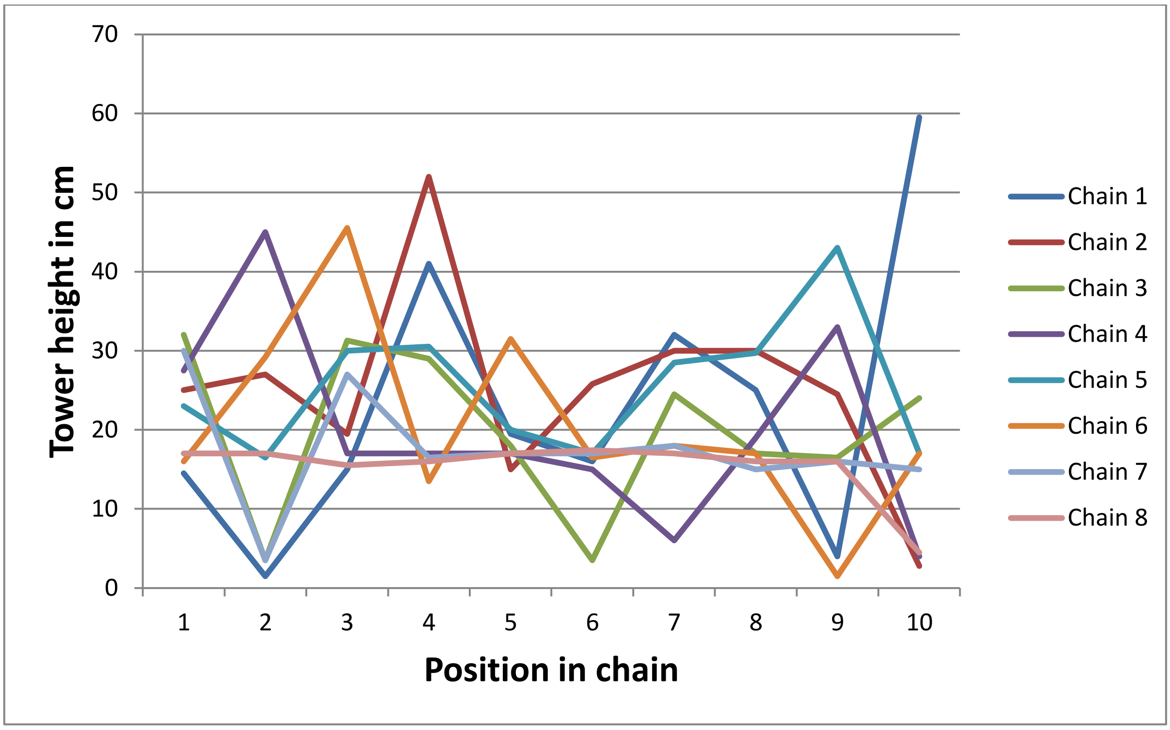


Fig. 11. Tower height across the 10 positions of the eight transmission chains.

5.3.2.2 Research question 2: Did children produce a culture-dependent technological product?

Since the previous analyses showed that there was no increase in tower height along the chains, we did not expect children at the end of the transmission chains to have outperformed children in the baseline. We compared the average tower height in position 10 across the chains with the average height in the baseline and indeed found no significant difference (independent samples t-test, two-tailed: *t*(27) = 0.150, *p* = .882, Cohen's *d* = 0.053). Similarly, there was no difference between the average tower height in positions 8 to 10 across the chains and the average height in the baseline (independent samples t-test, two-tailed: *t*(27) = -0.612, *p* = .545, Cohen's *d* = 0.272). With regard to tower shape, no child in the transmission chain condition produced a tripod. These results suggest that children in the transmission chains did not produce anything that went beyond what children in the baseline were able to produce independently. Thus, the children performed similarly to children in the baseline condition of this study as well as to the baseline conditions conducted in the studies of chapter 4. Whereas some of the children in the demonstration conditions of chapter 4, who were presented with an evidently good tower – the tripod – were readily able to copy the tripod, the current study suggests that seeing other children's towers, but no tripods, is insufficient for children to invent the tripod shape themselves. This might have been due to children's inability to innovate on each other's ideas or to invent the tripod shape on their own, or because children interpreted the testing situation in the studies in chapter 4 more strongly as a pedagogical context (they knew the tower were made by an adult, in contrast to the current study in which they were told that other children had made the towers), which might have made them more motivated to copy the demonstrated tower in the previous studies. However, we currently favour the idea that features of our current study design hindered children to show a basic capacity for the ratchet effect; this would imply that changes to the study design are promising future directions to eventually detect children's ability to produce a ratchet effect (see discussion).

5.3.2.3 Research question 3: Was there cultural variation between the transmission chains?

The mean similarity of the towers within their respective transmission chains was 1.80 (*SD* = 0.73), whereas the mean similarity of the towers to the towers of the remaining chains was 1.54 (*SD* = 0.49). Even though both ratings were relatively low, a two-tailed paired samples t-test revealed that the mean similarity *within* chains was significantly greater than the mean similarity *between* chains ($t(79) = 3.778$, $p < .001$, Cohen's $d = 0.422$), suggesting the presence of chain-specific design traditions (the towers made in the transmission chains are depicted in Appendix 6).

5.4 Discussion

Using the transmission chain paradigm, we investigated whether chains of children between 4.5 and 6 years were able to produce a culture-dependent technological product in a construction task that was adapted from the spaghetti tower task by Caldwell and Millen (2008a). Children were tested individually and asked to build something as tall as possible using plasticine and sticks. Participants in the transmission chain condition were provided with an endstate demonstration, i.e., they had the opportunity to see the constructions built by the two previously tested children (apart from children in generation 1, who were not presented with any construction, and those in generation 2, who only saw the construction made by the first child). Results showed that tower height did not increase along the chains. In addition, compared to the performance of baseline children, children in the transmission chains did not make towers that went beyond what children in the baseline were able to achieve, both with regard to tower shape and height: Children in the transmission chains did not produce a tripod-shaped tower (an efficient solution to the task) and neither did the average tower height in positions 8 through 10 in the transmission chains differ from the average tower height in the baseline, i.e., even after prolonged transmission chains, tower height did not improve over baseline performance. This is notwithstanding the fact that our treatment worked in that transmission within the chains did happen: We found evidence for within-chain design traditions, namely in that towers within their respective chains were more similar to each other than to towers of differing chains. Thus, we detected transmission, but we did not detect the ratchet effect; children in the transmission chains did not produce a culture-dependent technological product.

Despite children's ability to learn from and copy a culture-dependent technological product in the tower building task, even in the absence of action information (chapter 4), children do not seem able to produce a culture-dependent technological product by themselves, at least not when allocated to transmission chains in which they have access to the end-product of two previous generations. This finding contrasts with some studies on human adults showing that transmission chains of adult participants working on cognitively transparent tasks, i.e., tasks whose causal and physical characteristics can be rather easily retraced (e.g., building paper planes or simple baskets), are able to generate a ratchet effect even when – similar to our task – participants do not have access to the actions, but only to the end results of the previous generations (Caldwell & Millen, 2009; Zwirner & Thornton, 2015).

In the following we present several (non-exclusive) explanations for our finding. As cumulative culture has been argued to rest on two key cognitive abilities – high-fidelity transmission and innovation – we suggest possible explanations along these two aspects. Regarding transmission, one might think of the following explanations: First, it might be possible that children were indeed willing to use the social information provided to build better constructions themselves, but that they experienced *difficulties in recognizing good inventions* (e.g., due to a still limited understanding of the physical and causal aspects involved in the task). However, Want and Harris (2001) showed that children learn from action demonstrations even when they do not result in a correct solution, so one might argue that children would have benefitted from seeing other children's towers even when these children were not able to make very good towers. Another possibility is that children *required even more information* in order to learn from the demonstrations (e.g., a demonstration involving both actions and end states). However, we think this explanation is unlikely as in chapter 4 we showed that 4- to 6-year-olds *do* copy good inventions (the tripod) and do so without action demonstrations.

Another explanation relates to children's *lack of motivation* to learn from the demonstrated towers. It may have been the case that even though children would have been able to learn from and improve upon the towers in principle, they were simply not motivated to do so. This might have been because children might have preferred to make their own construction rather than a copy or because the demonstrated towers did not look good or "spectacular" enough. Alternatively, children might have been less motivated to learn from the towers because they knew that the towers were built by other children rather than (more knowledgeable) adults. In line with this suggestion, several studies have shown an age bias in children's social learning in that children preferentially learn from adults over peers when learning how to operate a puzzle box, learning novel object labels or simple game rules (see Wood et al., 2013). Another reason why children might not have been motivated to learn from the demonstrated towers could have been a lack of a direct incentive to look at the towers or to make the tower taller than the demonstrated ones, possibly because children were not explicitly told to make their towers taller than the ones shown to them. However, in chapter 4 we showed that children indeed learned from and even copied the demonstrated tripod despite there being no direct incentive or prompt to copy (apart from the tripod being an efficient tower). Nevertheless, note that in these studies children knew that the tripod was built by an adult and this might have been interpreted as a prompt for copying.

Regarding the aspect of innovation, we suggest two possible reasons for why children in the transmission chains did not show a ratchet effect: First, while children might have been motivated and cognitively able to learn from the demonstrated towers (assuming no limitation on the transmission aspect), it could have been possible that there was simply “nothing” to be picked up because good inventions occurred too rarely or in too subtle a way. In order to test for this *lack of “innovations from scratch”* (Charbonneau, 2015, p. 325) hypothesis, future studies could seed transmission chains with evidently good towers and examine what would happen to the seeded trait. If children would further improve upon the good initial invention, one could conclude that the lack of a ratchet effect in the current study was probably due to a lack of an invention that was worth learning from. However, if the seeded tower would not be improved upon by the children but just transmitted unchanged along the chains or if it disappeared from the chains, this would suggest that it is not the lack of a good initial invention that explains the lack of the ratchet effect in children; rather, this would point to cognitive or motivational factors preventing a ratchet effect. In addition, such seeded transmission chains will be able to test how salient an improved tower design has to be in order to be picked up by young children.

Alternatively, it might have been possible that children in our study were motivated and also able to learn from the demonstrated towers (again no limitation on the transmission side) and that there was indeed no lack of good inventions, but that children – once they copied (parts of the) demonstrated towers – were *unable to make “cumulative modifications”* (ibid.) on them. This could be due to children’s still developing innovative skills or due to a motivational issue (as there was no direct incentive for children to make further improvements). Indeed, several studies indicate that young children find it hard to invent novel solutions to a tool-making task or to invent novel strategies to extract even more rewards from a puzzle-box, suggesting that children's skills are still developing over the pre-school and primary school years (Beck et al., 2012; Carr et al., 2015; Chappell et al., 2013; Cutting et al., 2011; Nielsen et al., 2014; Sheridan et al., 2016; Whalley et al., 2017). Similarly, Hill (2010) argued that whereas innovations are mainly made by adults, children mainly copy and do not innovate themselves as much.

In a recent review, Carr et al. (2016) discuss possible prerequisites for innovations such as causal understanding of and insight into the task at hand, curiosity and explorative tendencies as well as creativity and divergent thinking (but see Beck, Williams, Cutting, Apperly, & Chappell, 2016, who found no link between children's divergent thinking and innovative abilities). With regard to causal understanding, we argued in chapter 4 that children in our studies had a still limited understanding of the physical or causal laws involved in the tower task, which might have made the task – and innovations therein – difficult for them. And indeed, Flynn, Turner, and Giraldeau (2016) hypothesised that children might be more likely to innovate on tasks they "believe are easy, or feel expert in" (p. 7). In order to investigate the role of causal understanding for children's ability to produce culture-dependent traits, future studies might employ activities that test for and/or enhance children's causal understanding of the tower task. Studies could also include measures for creativity and divergent thinking and even experimentally manipulate these factors in order to detect possible links with children's innovative skills in a ratcheting context.

Dean et al. (2012) suggested that when working together, i.e., when they have the opportunity to observe, communicate with, and teach others, children are able solve a task that requires “cumulative cultural learning” (p. 1117). Even though their study lacked a baseline and research currently being carried out in our lab suggests that children can solve this task individually (Gwilliams et al., 2017), their study shows that young children readily work together to solve problems. Muthukrishna and Henrich (2016) argued that innovation might not be so much a cognitive characteristic of individuals (suggesting the presence of “innovators” or “geniuses”) but that innovation is the result of combining previously isolated pieces of knowledge. This combination of information can take place within a single individual who happens to bring together different areas of knowledge, thus becoming a “nexus of previously isolated ideas” (ibid., p. 4). Yet, with regard to young children it might be unlikely that children become such a nexus given that they have not had much time yet to gather information about the world (and our task in particular). However, the combination of previously unconnected pieces of knowledge can also be brought about by interaction and collaboration of individuals in which new knowledge occurs via co-construction (*collaborative learning*, Tomasello et al., 1993). Indeed, a positive effect of such collective intelligence on performance has recently been demonstrated in non-human animals: Sasaki and Biro (2017) let three groups of pigeons repeatedly fly home (beeline 8.6 km). In the two control groups, pigeons were flying individually or in fixed pairs of groups. The experimental condition consisted of 10 groups with five successive generations of pairs of pigeons (i.e., a pair was allowed to fly together 12 times (constituting one generation) before one pigeon was replaced by a naïve pigeon and the next generation with 12 flights started). Although route efficiency improved in all conditions, the generations of pairs eventually outperformed both individual pigeons and fixed pairs, demonstrating the power of information pooling across several individuals.

An interesting point from this research is the authors' suggestion that the ratchet effect does not rely on "complex cognition" (p. 4) such as teaching and imitation and thus could also be found in other birds or insects. Even though their results and conclusions are currently still debated, if we follow Sasaki and Biro's argument, then we would have yet another reason to expect to find a ratchet effect among human children in future experiments. In general, we agree with the authors that more research is needed that integrates the study of cumulative cultural evolution and collective intelligence. Therefore, we suggest that future developmental research should explore the role of collaborative learning for the ratchet effect (e.g., by running transmission chains with two or more children per generation) in order to gain more insights into the ontogenetic origins of cumulative culture.

CHAPTER 6: GENERAL DISCUSSION

6.1 Summary

This thesis investigated the developmental origins of behaviours related to human material culture. Chapter 1 presented the theoretical background of the thesis: First, the ZLS theory and its assumption that human tool cultures develop on a basis of simple tool behaviours that every human can invent asocially. Second, I explained the notions of cumulative culture and the ratchet effect as well as Vygotsky's (1978) concept of the ZPD which form the background of the studies presented in chapters 4 and 5. I presented my view of how these three theoretical constructs dovetail and I worked out a new integrated theoretical framework (Fig. 3) that I hope will prove useful for the study of cultural behaviours on a species as well on an individual (psychological) level. I then highlighted the research gaps that this thesis aimed to address: identifying some of the tool-use behaviours lying within the human ZLS (chapters 2 and 3), determining when children become able to copy cumulative technological design and which social information they need to do so (chapter 4), and investigating whether groups of young children can already produce a ratchet effect (chapter 5).

Chapter 2 presented one of the first explicit latent solution tests on human tool use. In the literature there had only been few, but important studies on children's spontaneous tool-use abilities, such as Piaget's (1952) investigations into toddlers' ability to pull in out-of-reach objects, which had sparked several other studies looking into the onset and development of basic tool-using skills in children (e.g., E. Bates et al., 1980; Bechtel et al., 2013; Brown 1990; Deák, 2014; Chen et al., 2000; Keen, 2011; Rat-Fischer et al., 2012; Willatts, 1984). Our study is in line with this research tradition, but focuses not so much on when and how basic tool-use abilities emerge, but at whether a diverse range of slightly more complex tool-using skills can be invented independently by young children (yes/no). The GATTeB study represents the most recent addition to the study of human *unaided* tool-using skills (see also studies that look into the motor rather than cognitive origins of human tool use, e.g.,; Keen, Lee, & Adolph; Lockman, 2000, 2005; Kahrs, Jung & Lockman, 2012; 2013; Kahrs & Lockman, 2014), which complements the rich and research-intense study on children's *social learning* of tool use (e.g., Casler & Kelemen, 2005; Conolly & Dalgleish, 1989; Flynn & Whiten, 2010; Gardiner, Bjorklund, Greif, & Gray, 2012; McGuigan & Whiten, 2009). Our study explicitly follows the latent solution test approach suggested by Tennie et al. (2009); together with their "loop-making" study, our study is among the first in a series of latent solution tests in humans (see also our ATU study) and non-human animals (Bandini & Tennie, 2017, Bandini et al., in prep).

We identified 11 simple tool-use behaviours that young children can invent individually, i.e., behaviours that do not require social learning to be acquired and that thus form part of the ontogenetic basis of human tool use. By showing that humans, just as other great apes, possess latent solution tool behaviours, our findings also emphasize the continuity between human and great ape cognition. Since our tool tasks were based on behaviours shown by wild chimpanzees and orangutans, our research suggests that the last common ancestor of humans and great apes – living 14 million year ago – was probably also able to spontaneously produce these behaviours. Thus, we possibly also identified a phylogenetic basis of human tool use. However, while we have argued that the basic physical cognition skills in young children and great apes possibly lie on a similar level, we would also like to point towards studies that highlight that human tool use has a highly social nature from early on: Children from 1.5 years attribute specific functions to tools when presented with a tool-use demonstration and these attributions guide their tool-using behaviours (Casler, Eshleman, Greene, & Terziyan, 2011). And from preschool age, children develop stable tool category representations based on tool function (i.e., a certain tool is *for* a specific purpose; Casler & Kelemen, 2005).

Our study is also similar to the recent studies on children's tool innovation (Beck, et al., 2012; Carr et al., 2015; Chapell et al., 2013; Cutting et al., 2011; Nielsen et al., 2014; Sheridan et al., 2016; Whalley, Cutting, & Beck, 2017) as these studies also present individual children with tool-using problems, investigating whether children can make simple tools unaided. While I personally would view these studies as latent solution tests for hook-making, it is still debated whether hook-making is a human latent solution: as mentioned above, it is still unclear whether making a hook is something that children can invent on their own or whether they require previous experience with hooks in order to solve the hook-making task.

Lastly, our study contributes to the literature the test battery created for this study (GATTeB), which hopefully can be used in future research. Indeed, researchers have already started to use it for different studies (Hoicka, pers. comm.; Neldner & Nielsen, pers. comm.). Due to its non-verbal nature, the GATTeB could also be adapted for non-human subjects.

Chapter 3 extended the latent solution test approach to the study of ATU. To my knowledge, it presents the first study so far examining different ATU types in children. We found that young children spontaneously use two tools in different combinations (tool set, sequential tool use, metatool use) in order to solve problems. Since the tasks were – where possible – closely modelled to spontaneous behaviours reported in wild or captive animals, our study allows for cross-species comparisons. Our study adds a previously overlooked comparison from the human side to the growing literature on spontaneous ATU capacities in great apes and birds (Köhler, 1925; Taylor et al., 2007, 2010; Martin-Ordas et al., 2012; Mulcahy et al., 2005; Wimpenny et al., 2009). It shows that human children, too, can invent several ATU behaviours on their own, without requiring social learning. However, it needs to be noted that in comparison to the studies on non-human animals, some of our task features were relatively simple (e.g., the positioning of the tools in the sequential tool use tasks), so there is scope for future research investigating (slightly) more complex ATU problems in young children. Our study adds to the list of latent tool behaviours identified in chapter 2 the variations of sequential tool use, tool set, and metatool use studied in chapter 3. Again, the tasks created for our study can be used for future research and are suitable for investigating spontaneous ATU abilities in other species.

As we also mentioned above, our tasks could also have been analysed with regard to investigating children's problem-solving and planning skills. In fact, the accounts by Alpert (1928) and Matheson (1931) provide such analyses in that they report several detailed case studies (following Köhler's, 1925, approach) to illustrate children's problem-solving behaviour, typical mistakes that they make and factors that enable or prevent task solution (e.g., visual attention, chance, insight). Our study differs from this approach in that we simply looked at whether certain ATU were absent or present in children; however, future research could analyse our data in a similar way. By that, our study could not only give insight in children's tool-use abilities, but could also add to the existing literature on children's problem-solving and planning (see e.g., Tecwyn, Thorpe, & Chappell, 2013, 2014; Völter & Call, 2014).

In chapter 4 the focus of the thesis shifted from latent solution tests to the study of children's social learning in the context of material culture. A key feature of human culture is our capacity for high-fidelity transmission of cumulative culture, i.e., of traits that are too complex or unlikely to be (re-)invented individually (Boyd & Richerson, 1996). Chapter 4 investigated the developmental origins of our ability to acquire culture-dependent traits. For this, we successfully adapted the "spaghetti tower task" (Caldwell & Millen, 2008a), demonstrating that our task can be used to study the acquisition of material culture-dependent traits in young children. Prior to our study, the research that had investigated whether young children can copy something that they themselves were unable or highly unlikely to invent was restricted to studies on social-conventional culture. That is, young children were demonstrated with novel, arbitrary actions such as switching a light on with the forehead (Gergely et al., 2002) or opening the lid of a box with a tool rather than with the hand (Nielsen, 2006). These studies, along with our knowledge about language acquisition, show that children acquire culture-dependent traits from their second year of life. Our study now complements this research in the domain of material culture. By nature of our task, we chose an older age group of children as children need to bring a certain level of motor skills, physical knowledge, and causal understanding to the task. Future research might be able to identify tasks which allow the study of the acquisition of culture-dependent material products at an even younger age.

Our findings showed, in line with the findings of some adult studies (Caldwell & Millen, 2009; Caldwell et al., 2012), that young children were able to copy cumulative technological design even in the absence of action information, i.e., via emulation. The study thus contributes another piece of evidence to the current debate of whether action-copying is always required for material cumulative cultural traits to be transmitted or whether copying results via one's own behavioural means (emulation) is faithful enough. As Heyes (2016b) stated, it might be possible that researchers so far have overestimated the role of imitation for high-fidelity transmission and at the same time neglected the power of other social learning mechanisms such as emulation. However, a final answer to the question which role emulation plays for material cumulative culture will require more studies using a variety of tasks of differing complexity.

An interesting connection can be made between our study and Legare and Nielsen's (2015) consideration of the different cognitive demands of social-conventional and instrumental skills. The authors refer to recent evidence showing that when interpreting a demonstration as conventional, learners pay more attention to the process and thus employ more imitation than is the case when they view a demonstration as instrumental, where they tend to focus on the endstate (Clegg & Legare, 2016). This claim could be further investigated by using our tower construction task and presenting it in different contexts (conventional, instrumental) to children. In addition, Legare and Nielsen (2015) suggest that the rate of innovation in instrumental tasks should be greater than in social-conventional tasks (see also Legare, Wen, Herrmann, & Whitehouse, 2015). However, independent studies are still lacking. It would be interesting to use our tower task to further study children's innovative capacities in an instrumental setting. As Caldwell, Cornish, and Kandler (2016) argued, one might expect to find a higher rate of innovation in tasks in which participants are given a performance goal (i.e., they asked to achieve the best performance possible, e.g. build something as tall as possible) compared to a repetition goal (i.e., when asked to explicitly copy the demonstrated trait).

One might ask in how far our study provides novel insights, given that we already know that children are cultural magnets (Flynn, 2008) and possess a portfolio of social learning mechanisms to acquire cultural traits from early on (Whiten et al., 2009; see also McGuigan, Burdett, Burgess, Dean, Lucas, Vale, & Whiten, in press). The crucial difference to previous social learning studies is that, in contrast to these earlier studies (e.g., Hopper et al., 2010; Nielsen, 2006), our task did not provide children with a demonstration of simple actions that children in a baseline could readily invent on their own. Implementing a baseline allowed us to provide children with a demonstration of a culture-dependent product.

Having said this, we need to emphasize that the studies presented in chapter 4 were not designed to study "cumulative culture in children". As of recently, researchers have set out to study the roots of cumulative cultural learning (the ratchet effect) in children (Dean et al., 2012; Tennie et al., 2014; and most recently McGuigan et al., in press), as we have done in chapter 5. However, with the studies in chapter 4 we did not look at the ratchet effect but explicitly "went one step back" in order to investigate whether children possess one of the two key cognitive requirements for producing a ratchet effect – high-fidelity learning of a cumulative cultural product (the other key component is innovation). We showed that this is the case and provide important empirical findings for the further study of the ratchet effect in children.

Having shown that young children were able to copy cumulative technological design, chapter 5 went on to investigate whether groups of children would also be able to produce cumulative technological design on their own, i.e., whether there was evidence of a ratchet effect. For this, we adapted the transmission chain paradigm used in studies on human adults to the case of children, using our tower task (Caldwell & Millen, 2009).

Our study contributes to the literature in that it presents one of only few studies attempting to investigate the additive ratchet effect in children, i.e., whether groups of children can produce a cultural trait that is beyond what children can achieve individually. However, in contrast to previous findings with adults, we did not find evidence for a ratchet effect. We presented a series of cognitive and motivational explanations for this result and have suggested ideas for future research.

Very recently, a study by McGuigan et al. (in press) employed an open-diffusion design to study the ratchet effect in children in a complex puzzle-box environment consisting of four levels. As in Dean et al. (2012), the levels contained rewards of increasing desirability. The authors found that groups of children were able to reach stages that asocially tested baseline children did not reach on their own. By identifying cycles of inventions, social learning of these and their subsequent spread through the group, this study contributes important insights into the roots of cumulative cultural learning in children. However, the results also need to be interpreted with caution, as the duration of the asocial control condition was rather short (45 min baseline vs 4h main condition), possibly masking the children's ability to reach higher stages on their own. In addition, not all control conditions were realised ("L3-only" control was missing) and evidence that social learning was needed to reach level 4 remained circumstantial. In contrast to McGuigan et al. (in press) as well as other previous studies (Dean et al., 2012; Tennie et al., 2014), our task design is open-ended and thus allows for a potential ratchet effect to occur. In addition, it includes a baseline condition which previous studies lacked (e.g., Dean et al., 2012) and which we treated as a central rather than subsidiary element of our study.

On a more theoretical level, the chapter suggested some refined terminology that might prove useful when describing the study of "cumulative culture in children".

6.2 Limitations of the current work

"[T]he experimenter should recognize that every intelligence test is a test, not only of the creature examined, but also of the experimenter himself." (Köhler, 1925, p. 265)

Next, I will address the limitations and possible criticisms of the studies presented here. Starting with chapter 2, it has to be noted that the GATTeB study has not been independently replicated yet. As mentioned above, reproduction of novel findings is crucial to enhance their credibility, especially when a study is the first of its kind (Open Science Collaboration, 2015)[20]. In addition, one might criticize that the GATTeB mainly consists of stick tool-use tasks (all tasks apart from nutcracking involve the use of a stick) and that this does not reflect the breadth of the great ape (and potentially child) spontaneous tool behaviours. We agree with this point, which was raised by one of the reviewers of the paper reporting the GATTeB results, but need to point out that we were not allowed to take certain material (leaves) into nurseries and so we had to deselect several potentially interesting tasks. Early exploratory work in which we did present children with some leaves (not reported here) suggested that children readily picked the

[20] However, a replication study with a Western and non-Western sample is on its way (Neldner & Nielsen, pers. comm.).

leaves up and also used them to wipe up liquids. Future studies, carried out in different locations, could add “leaf tasks” to the GATTeB.

Another criticism was raised by a researcher studying great apes in the wild: The GATTeB tasks we presented to children might not be comparable to wild great ape behaviours because in the wild, a given tool-use behaviour almost never just consists of the sole use of the tool – e.g., termite-fishing is never just “inserting a stick into a hole to get insects”. Rather, so it is argued, the tool behaviours consist of many preceding steps, e.g., in the case of termite-fishing it also entails searching for, identifying, and potentially modifying a suitable stick as well as using one’s knowledge about the environment, e.g., with regard to the location of termite mounds. All these parts of the whole act need to be learned by the chimpanzee which will likely require more than just a single tool-use instance and might be facilitated by the use of social information. In contrast to the case in wild great apes, it is argued, our study presented children with only the final step of the behavioural sequence (e.g., to insert the stick into a hole) and already provided children with all the necessary (and only the necessary) raw materials. Therefore, it has been criticized that our tasks are too easy, potentially overestimating children’s spontaneous tool-using capacities and thus not adequately representing the challenges of tool use in wild great apes. In reply to this, we would like to emphasize the basic approach suggested by the ZLS theory, namely that latent solution tests should test for the spontaneous invention of those behaviours for which a “cultural claim” (Reindl et al., in press, p. 8) has been made, e.g., by the respective reviews of potentially cultural behaviours in chimpanzees (Boesch, 2012; Whiten et al., 1999, 2001), orangutans (van Schaik et al., 2003, 2009), and – more recently – in gorillas (Robbins et al., 2016). In Reindl et al. (in press), we have exemplified this logic:

For instance, if the behavior involves a long walk to gather a stick prior to using it in a special, supposedly "culturally learned" way, then the latent solution test does not need to recreate the "long walk" aspect. What matters is that the stick is used by naïve subjects in the same way as in the target population. (p. 8)

Somewhat related to this point, I would like to emphasize that we do not argue that social learning plays no role for the acquisition of tool behaviours of great apes. As the ZLS theory hypothesises (Tennie et al., 2009) and as latent solution tests on captive apes carried out in our lab show (Bandini & Tennie, 2017, Bandini et al., in prep.; Tennie, Hedwig, Call, & Tomasello, 2008), great apes do not *require* social learning in order to learn certain tool behaviours. Again, this does not rule out the possibility that social learning has an influence on the likelihood that this individual invention will occur in a given instance. Thus, we do not go against the notion of social learning playing a role for the prevalence of a given behaviour in a great ape community.

Moving to chapter 3 (ATU), here, too, I would like to point out the need for replication. Again, this is because of a general need to replicate first-time results. In addition, there is the need to introduce design changes to the Anvil prop task which in our study was ill-designed (the task did not make the need for a metatool sufficiently salient and it allowed for alternative solutions). Lastly, future studies are needed to follow up on the question of whether there is a hierarchy in difficulty between the ATU types (in order of increasing difficulty, as suggested by our findings: tool set – metatool use – sequential tool use).

With regard to chapter 4, we acknowledge that the claim that the tripod represents cumulative technological design for children aged 4 to 6 years is based on the assumption that children in baseline conditions do not spontaneously build a tripod. If future studies should find children spontaneously making tripods, the claim of our study that the tripod represents cumulative technological design for 4- to 6-year-olds would have to be abandoned. However, having conducted the baseline condition in two separate studies and having chosen a sufficiently young age range, we are quite confident that the tripod shape is just beyond what children between 4 and 6 can invent on their own.

Finally, two critical points have to be made about the transmission chain study in chapter 5. First, the fact that we encouraged children to use the full building time (10 min) resulted in several children disassembling their fairly tall towers in order to rebuild another one (that sometimes ended up being smaller than the previous tower); in other cases, children's initial towers crashed because they tried to make them even taller as a response to our encouragement. As a result, several children in our study produced as their final tower a tower that did not also represent the tallest tower they made. Thus, in these cases, the towers that were presented to the next "generation" were not representatives of the best performance achieved by the previous generation. Therefore, we might have reduced the opportunities for children to detect good towers among the demonstrated ones and thus limited the opportunities for a ratchet effect to occur. When trying to identify reasons why our study did not find a ratchet effect in children, future studies need to introduce changes to the study design in order to address this point. For example, the stopping rules could be changed so that trials would end when children indicate they are finished rather than when the maximum time is over. Alternatively, one could allow children to build several towers within the 10 min timeframe rather than granting them only one building board that potentially encouraged the disassembly of towers. Then, E could choose the tallest tower at trial end to be presented to the next generation.

Another, more general criticism of the study relates to the nature of the transmission chain methodology. It might be criticized that cultural transmission “in real life” rarely looks like the “telephone game” chain employed in our study (one individual per generation, one transmission event, no side branches). Rather, as Morin (2016a, p. 125) described, “transmission in real settings can be repeated”, “can come from several distinct individuals, and not just one model”, and a “single diffusion chain can, at any point, branch out into multiple chains”. Thus, our study design could be criticized for making it especially difficult for innovations to accumulate. Another factor contributing to this difficulty might be the fact that our design did not allow for transmission through imitation, but restricted children to emulation learning. In addition, this procedure might also have reduced the ecological validity of the study. It is evident that many more studies need to be conducted in this field (both with adults and children) and we hope that the current study has contributed by pointing towards possible future pathways.

6.3 Suggestions for future work

“The more I learn, the more I realize how much I don't know.” (Albert Einstein)

I hope that the findings presented in this thesis have raised new questions and will stir ideas about new pathways for research. For example, in order to support our claim that the GATTeB study identified tool-use behaviours representing latent solutions in humans, future studies need to replicate our work in different, non-Western countries so that we can base our conclusions on evidence from a wider sample of human populations (Henrich, Heine, & Norenzayan, 2010; Legare & Harris, 2016). We are excited about such kind of research already being under way (Neldner & Nielsen, pers. comm.).

Future studies could also take a completely different approach to identifying potential human latent solutions: The idea here is to identify types of tool use that occur in many different cultural groups of humans; this cross-cultural “baseline” of tool behaviours could hint towards tool behaviours which are relatively easy to invent. These behaviours could then be adapted to tasks for human children and tested in latent solution experiments as done in chapter 2. Specifically, since modern human groups around the world are characterised by possessing more or less large repertoires of cumulative culture, we suggest studying the tool repertoires of many small, unconnected, relatively isolated hunter-gatherer societies before they had been affected by modern European culture (this could be done by using the reports collected in anthropological databases such as the electronic Human Relation Area Files (eHRAF, ehrafworldcultures.yale.edu)). One would only choose groups which are judged to resemble Pleistocene hunter-gatherers, i.e., anatomically modern humans living between 75 000 years before 40 000 years ago (Boehm, n.d.). This is because it is assumed that Pleistocene hunter-gatherers were not yet as inventive and culturally flexible as modern humans (maybe due to restrictions in their population densities; ibid.) and therefore their capacities for producing and maintaining complex cumulative cultural tools are likely to be reduced (see also Henrich, 2004). This argumentation is debatable, of course, but such an approach might still be helpful to identify further possible candidates for human latent solutions in the tool-use domain. It is important to say that we would not conclude from a finding that a certain tool behaviour was found in every of the studies hunter-gatherer groups that this behaviour constituted a latent solution; rather, this behaviour would then be adapted to be a latent solution test for children, and only this test would provide first answers.

A different pathway for future research regards the issue of comparing the ZLS between humans and great apes. As mentioned in chapter 2, it is still an open question whether the range of latent tool-use solutions in humans and great apes is comparable or whether the human tool ZLS proves to be larger. To be able to answer this, a new test battery of tool tasks would need to be created, this time consisting of tasks that are neither based on great ape tool behaviour nor inspired from the human side. Only those tasks that are equally novel to human children and great apes are able to answer the question whether the spontaneous tool-use capacities in human children and great apes are on a similar level or whether children actually outperform our closest living relatives. A potential starting point for finding such tasks could be to search for suitable tool behaviours shown by animals outside of the great ape lineage (e.g., monkeys or birds) because these would arguably be equally unfamiliar to both children and great apes.

Moving to our work on children's social learning, we have suggested that future studies should further examine the role of emulation for the transmission of cumulative culture. In addition, as we pointed out above, future work also needs to investigate the interaction between the social learning mechanism required to faithfully transmit cumulative culture and the cognitive transparency of the cultural trait in question. Having shown that children are able to copy cumulative technological design via reverse-engineering, I think that future research should also address more the role of reconstructive (as opposed to replicative) processes in the transmission of culture-dependent traits. This is especially important as proponents of the cultural attraction theory claim that the phenomenon of reconstruction has been under-researched and underrepresented (Morin, 2016a; Scott-Phillips, 2017; Sperber, 1996). According to Sperber (1996), cultural transmission is not a copying process of mental concepts. Rather, in order to be transmitted, cultural traits first need to be expressed – e.g., in the form of an action, a verbal explanation or a sketch – and this expression in turn is the basis for a reconstructive process in the learner. Even in cases where the demonstrator and the learner eventually possess the same cultural trait the traits are not identical as the trait is always reconstructed, not copied. While the claim that reconstruction has been neglected in the literature is currently much debated (Acerbi & Mesoudi, 2015; Morin, 2016b), it is evident that more efforts have to be made to find a common language between representatives of different schools of thought. For example, reconstruction (i.e., at least partly re-building a trait rather than replicating it) is claimed to be part of every instance of cultural transmission (i.e., imitation and emulation; Sperber, 1996). However, it seems unclear how the term reconstruction relates to the concepts of imitation and emulation. For example, when learning through emulation, a learner uses information about a demonstrator's goal and the results of her actions; since action information is lacking, the learner has to achieve the same result via her own behavioural means ("reverse-engineer"). Are the results of this action a replication or a reconstruction of the demonstrated results? When asked about his view on the conceptual relationship between reconstruction and emulation, Morin (pers. comm.) stated that he did not use the term emulation at all. Clearly, more efforts have to be made by all the researchers in the field to clarify concepts and how they relate to each other to facilitate communication. There are already promising attempts to reconcile standpoints (Buskell, 2017; Sterelny, 2017) but it is clear that more research and exchange needs to take place in order to fully benefit from the contributions of different theoretical approaches to the study of human culture.

With regard to our work on the ratchet effect in children, it has to be noted that the tower building task cannot be adapted for non-human animals because it requires verbal instructions and because it is not directly related to obtaining rewards. Therefore, more work is needed that creates tasks that can be used with several species (e.g., Dean et al., 2012).

One of the most interesting questions from the work in chapter 5 is under which conditions a ratchet effect can be evoked in children. An intriguing aspect to look at could be the role of collaboration in the production of culture-dependent traits. As Vygotsky (1978) and Tomasello et al. (1993) have outlined, children at a similar developmental level are able – via collaborative learning – to co-construct new knowledge that goes beyond what each individual child could have reached independently. This process of "cultural creation or coconstruction" (Tomasello et al., 1993, p. 501) thus carries the seeds of the creation of culture-dependent behaviours in children. For collaboration and interaction represent important components for humans' ability to produce cumulative culture as cumulative culture has been argued to be the product of the collective brain (Muthukrishna & Henrich, 2016). Therefore, I think that investigating the collaborative nature of social learning in children can tell us more about the deep developmental roots of human cumulative culture.

This thesis focused on the domain of material culture and so did not discuss what we (do not yet) know about the development of social-conventional culture. Clearly, work in both domains is needed to obtain a comprehensive picture of the ontogenetic development of human culture. Research on children's learning of social conventions is still comparatively sparse (see e.g., P. L. Harris, 2012; Legare & Nielsen, 2015; Nielsen, Kapitány, & Elkins, 2015; Rybanska, McKay, Jong, & Whitehouse, 2017) and so presents another promising domain for future studies. As Legare and Nielsen (2015) pointed out, there is no strict distinction between instrumental and social-conventional behaviours; rather, human behaviour is often “a blend of instrumental and conventional acts” (p. 692). Therefore, these authors have started to discuss the characteristics and relationships of instrumental and social conventional learning and to directly compare children’s social learning in instrumental and social-conventional contexts (see studies in Legare & Nielsen, 2015; Wilks, Kapitány & Nielsen, 2016). These studies have also raised many new questions that need to be targeted in the future.

Here I presented just some of many possible pathways of prospective research. Future studies on children’s instrumental and conventional social learning, the social learning mechanisms and biases as well as children’s collaborative and innovative capacities will yield many more fascinating and comprehensive insights into the development of human culture that will hopefully be of interest to many disciplines studying culture.

APPENDICES

Appendix 1: The zone of actual development (ZAD) and the zone of latent solutions (ZLS) applied to the case of non-human great apes.

The figure below depicts how the ZAD and the ZLS intertwine in the case of non-human great apes. According to Vygotsky (1978), chimpanzees do not possess a ZPD as they can only learn what they could have invented themselves. Consequently, the great ape ZAD only consists of ZLS behaviours. On the species level, the ZLS theory states that great apes are not able to produce cumulative culture (CC; Tennie et al., 2009). Therefore, a ZPD and CC are missing in this figure. Note that I also assumed a ZLS_{far} for great apes (i.e., the presence of latent solutions which are not within the ZAD from birth; light blue area on the left). However, given the assumption of the ZLS theory (as well as current evidence, see chapter 1) that great apes do not learn through high-fidelity social learning, I do not assume that behaviours within the ZLS_{far} can be acquired through high-fidelity social learning (which I assume for the human case, see chapter 1). This is because great apes are assumed to learn traits only either individually or through a combination of low-fidelity social learning (e.g., stimulus enhancement) and individual learning (Tennie et al., 2009). Fig. A1-1 thus presents the ZLS-only claim for great apes. If future research showed that great apes copy behaviours lying outside their ZLS, i.e., if it showed that great apes did have a ZPD, the statements of the ZLS theory and this figure would have to be adjusted accordingly. In that case, great apes would be placed into a similar category as humans: “into one where the ZLS potentially explains only part of their behavio[u]ral repertoire” (ZLS-plus claim; Reindl et al., in press).

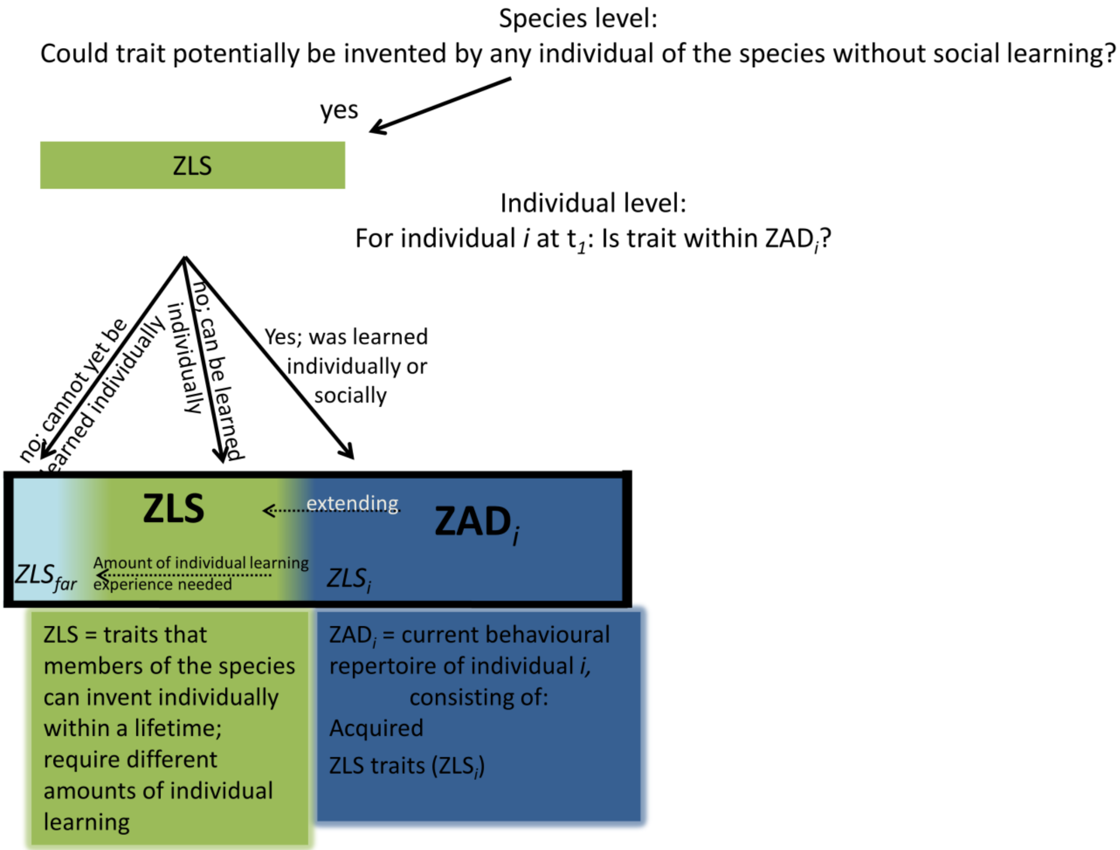


Fig. A1-1. The zone of actual development (ZAD) and the zone of latent solutions (ZLS) applied to the case of non-human great apes.

Appendix 2: GATTeB instructions

English version

Warm-up task

Children were presented with the picture of the meadow and the farmer and asked whether they could help the farmer build a fence. Children were then shown a stick of Balsa wood and encouraged to break it into smaller pieces. If children were hesitating, E repeated the encouragements or helped them complete the task.

E said "Before we start, I'd like to show you something! Look, here is a large meadow with some animals on it! Do you know these animals? ...What kind of animals are these? ...Yes, horses! How many horses can you see? Let's count them! ...The horses belong to farmer Joe! Here he is! Joe wants to build a fence so that his horses cannot run away. He has already begun building the fence on this side. Do you think you can help Joe finish the fence?...Here is a long stick and you can break this stick into smaller pieces like this one so that we can make the fence. Do you think you can do that?"

Insect-pound (IN)

E put the apparatus and the tool in front of the child. She then presented three balls of Play Doh and put them into the tube. Children were told that they had to retrieve the balls in order to win a sticker. The tiny spikes at the end of the stick allowed the Play Doh to stick to the tool. To count as correct tool use, children had to insert the stick into the tube and to pound the balls at the bottom of the tube. Correct success was scored when at least one of the balls was retrieved from the tube. Children had 2 min to complete this task.

E said “This is our first/next game. Here is a tube and here I have three balls of Play Doh. Look what I do! One, two, three! If you can get the balls out again, you win a sticker!”

Perforate (PER)

E presented the child with the apparatus and the tool and drew his or her attention to the die with the attached sticker in the box. Children were told that they could keep the sticker if there were able to retrieve it from the box. Correct tool use was scored when children inserted the stick into the apparatus and broke the barrier of flower arrangement foam. Correct success was scored when children removed the die from the box by tipping the apparatus. Children had 2 min to complete this task.

E said “This is our first/next game. Look, there is a sticker in the box! If you can get the sticker out of there, you can keep it!”

Nuthammer (NUT)

E put the apparatus, the plastic nut, and the clay hammer in front of the child and told him or her that the game was to open the ball to retrieve a sticker. Correct tool use was scored when children took the hammer in one or two hands and hit it onto the nut. Correct success was scored when children opened the plastic nut after using the hammer. Children had 2 min to complete this task.

E said “This is our first/next game. Here is a ball and there is a sticker in it. If you can get the sticker out of there, you can keep it!”

Algae scoop (AE)

E put the apparatus and the tool in front of the child and drew his or her attention to the sticker attached to a strip of plastic inside the box. Children were told that they could keep the sticker if they were able to retrieve it from the apparatus. Correct tool use was scored when children inserted the stick into the box and touched the strip of plastic. Correct success was scored when the strip of plastic was fully removed from the box. Children had 2 min to complete this task.

E said "This is our first/next game. Can you see the sticker? If you can get the sticker out of there, you can keep it!"

Ground puncture (GR)

Due to its size, the apparatus in this game was placed on the floor. If children were sitting on a chair, E turned the chair in the direction of the apparatus. E held the box up and drew the child's attention to the sticker which could be seen through one of the windows. Children were told they could keep the sticker if they were able to remove it from the box. Correct tool use was scored when children used the stick to make a hole in the plasticine or at least tried to puncture the plasticine layer. Correct success was scored when children made a hole and removed the sticker by putting their hand through the hole. Children had 3 min to complete this task.

E said "This is our first/next game. There's a sticker in the box, can you see it? If you can get the sticker out of there, you can keep it!"

Seed extraction/Nut extract (SEED)

E put the apparatus and the tool in front of the child and asked whether he or she could see the pom poms inside the box. Children were then told they could win a sticker if they were able to retrieve one of the balls. Children had to use the stick in a levering fashion to push the pom poms through the small slit on top of the box. Correct tool use was scored when children inserted the tool into the apparatus and used it to lever the pom poms out of the box. Correct success was scored when at least one of the balls was retrieved from the apparatus. Children had 2 min to complete this task.

E said "This is our first/next game. There are colourful balls in this box, can you see them? If you can get them out of there, you win a sticker!"

Marrow pick (MA)

E put the apparatus and the tool in front of the child and drew his or her attention to the sticker attached to the sponge inside the tube. Children were told they could keep the sticker if they could retrieve it from the tube. Correct tool use was scored when children inserted the stick into the tube and touched the sponge with it. Correct success was scored when children fully retrieved the sponge from the tube. Children had 1 min to complete this task.

E said "This is our first/next game. There is a sponge with a sticker in the tube. Can you see it? If you can get the sticker out of there, you can keep it!"

Fluid-dip (FD)

Children were presented with the apparatus, the bottle lid, and the stick and were asked to have a look into the tube. They were told that there was yellow paint inside the tube and that they could win a sticker if they could get some of the paint out of the tube and into the container. Correct tool use was scored when children inserted the stick into the tube and correct success was scored when they were able to place some paint into the container. Children had 1 min to complete the task.

E said "This is our first/next game. In this box, there is something yellow! Can you see it? It is yellow paint! If you can get a little bit of the paint out of there and into here, you win a sticker!"

Ant-dip-wipe (ADW)

Because of its size, the apparatus was placed on the floor. If the child was sitting on a chair, E turned the chair towards the apparatus. The stick was already inserted in the box. Children were asked whether they could see the white balls inside the box and were told they could win a sticker if they could get some of the balls out of the box and into another box next to the apparatus. Since the focus of this task was not on whether children spontaneously used the tool per se, but on whether they used a certain efficient strategy to remove the balls from the stick, children were encouraged to pull the stick from the apparatus. Correct tool use was scored when children held the stick in one hand while wiping off the balls with the other hand – either using a close grip, the flat hand or the finger tips. Correct success was scored when children were able to remove all the balls from the stick with one of the wiping behaviours. Children had 3 min to complete this task.

E said "This is our first/next game. Can you see the white balls in the box? Do you think you can get some of these balls out of there and into this box? If you can do that, you win a sticker! ...Try to pull this stick out the box and see what happens!"

Termite-fish leaf-midrib (TFLF)

E put the apparatus and the tool in front of the child. She then presented three small pieces of sponge scourer with stars glued to them and put them into the box. Children were told they would win a sticker if they could retrieve the stars from the apparatus. The stick had Velcro at both ends so that the stars could easily get attached to it. Attached to the stick was a paper leaf in such a way that it was impossible to reach the stars without tearing the leaf off the stick first; either end of the stick was too short to reach the stars when the leaf was still attached. Correct tool use was scored when children inserted the stick into the box after they tore the leaf off the stick either by ripping it off with one hand or by pushing the stick into the box. Correct success was scored when children retrieved the stars with the stick after ripping off the leaf. Children had 2 min to complete this task.

E said "This is our first/next game. Here is a box and here I have three stars. Look what I do! One, two, three! If you can get the stars out of there, you can win a sticker!"

Lever open/stick as chisel (LEV)

E put the apparatus and the tool in front of the children. She then drew children's attention to the ball in the mug by shaking the apparatus and telling children that there was a ball inside. Children were told they could win a sticker if they were able to get the ball out of the mug. The plasticine lid of the mug already contained a small hole which children were supposed to enlarge with the tool. E pointed the hole out to the children and asked them whether they could make it larger. Correct tool use was scored when children either inserted the stick into this small hole to make it wider or when they tried to make a new hole in another place on the lid. Correct success was scored when children retrieved the ball from the mug by tipping the apparatus after they made a big hole into the lid. Children had 1 min to complete this task.

E said “This is our first/next game. There is a ball in the mug, can you hear it? There is a sticker on the ball. If you can get the ball out of there, you can keep the sticker! Look, there is already a hole in the lid. Do you think you can make it larger?”

Termite-fish/Tree-hole tool-use (TF)

E put the apparatus and the tool in front of the child. She then presented three small pieces of sponge scourer with stars glued to them and put them into the box. Children were told they would win a sticker if they could retrieve the stars from the apparatus. The stick had Velcro at both ends so that the stars could easily get attached to the tool. Correct tool use was scored when children inserted the stick into the box. Correct success was scored when children retrieved at least one star by using the tool. Children had 1 min to complete this task.

E said “This is our first/next game. Here is a box and here I have three stars. Look what I do! One, two, three! If you can get the stars out of there, you can win a sticker!”

English version

Warm-up task

E said "Bevor wir anfangen, möchte ich dir noch etwas zeigen! Guck mal, hier ist eine große Wiese mit Tieren. Kennst du die Tiere?...Welche Tiere sind denn das?...Genau, Pferde!Wie viele Pferde kannst du sehen? Komm, wir zählen sie!...Die Pferde gehören Bauer Joe! Da ist er! Joe will einen Zaun bauen, damit seine Pferde nicht davonlaufen können. Auf dieser Seite hat er schon angefangen. Denkst du, du kannst Joe helfen, den Zaun fertig zu bauen?...Da ist ein langer Stab und den kannst du in kleine Stücke so wie das hier brechen, dann können wir den Zaun bauen! Denkst du, du kannst das machen?"

Insect-pound

E said "Das ist unser erstes/nächstes Spiel. Hier ist eine Röhre und da habe ich ein paar Kugeln aus Play Doh! Schau, was ich mache! Eins, zwei, drei! Wenn du die Kugeln wieder rausholen kannst, gewinnst du einen Sticker!"

Perforate

E said "Das ist unser erstes/nächstes Spiel. Guck mal, da ist ein Sticker in der Box! Wenn du den Sticker da rausholen kannst, darfst du ihn behalten!"

Nuthammer

E said "Das ist unser erstes/nächstes Spiel. Hier habe ich eine kleine Kugel und da ist ein Sticker drin. Wenn du den Sticker da rausholen kannst, darfst du ihn behalten!"

Algae scoop

E said “Das ist unser erstes/nächstes Spiel. Kannst du den Sticker sehen? Wenn du den Sticker da rausholen kannst, kannst du ihn behalten!“

Ground puncture

E said “Das ist unser erstes/nächstes Spiel. In der Kiste ist ein Sticker, kannst du ihn sehen? Wenn du den Sticker da rausholen kannst, kannst du ihn behalten!“

Seed extraction/Nut extract

E said “Das ist unser erstes/nächstes Spiel. In dieser Box sind bunte Bälle, kannst du sie sehen? Wenn du sie da rausholen kannst, gewinnst du einen Sticker!“

Marrow pick

E said “Das ist unser erstes/nächstes Spiel. Da ist ein Schwamm mit einem Sticker in der Röhre. Kannst du ihn sehen? Wenn du den Sticker rausholen kannst, darfst du ihn behalten!“

Fluid dip

E said “Das ist unser erstes/nächstes Spiel. In dieser Box ist etwas Gelbes! Kannst du es sehen? Das ist gelbe Farbe! Wenn du ein bisschen Farbe da rausholen und da rein tun kannst, gewinnst du einen Sticker!“

Ant-dip-wipe

E said “Das ist unser erstes/nächstes Spiel. Kannst du die weißen Bälle in der Box sehen? Glaubst du, du kannst ein paar von den Bällen da rausholen und in diese Kiste tun?

Wenn du das tun kannst, gewinnst du einen Sticker! Versuch mal, den Stab aus der Box zu ziehen und guck mal, was passiert!“

Termite-fish leaf-midrib

E said “Das ist unser erstes/nächstes Spiel. Hier ist eine Box und da sind drei Sterne. Schau, was ich mache! Eins, zwei, drei! Wenn du die Sterne wieder rausholen kannst, gewinnst du einen Sticker!“

Lever open/stick as chisel

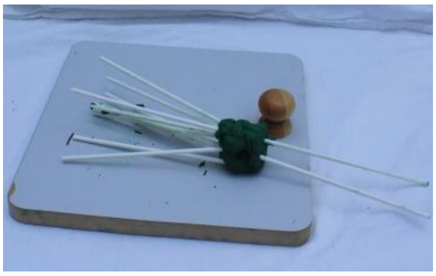
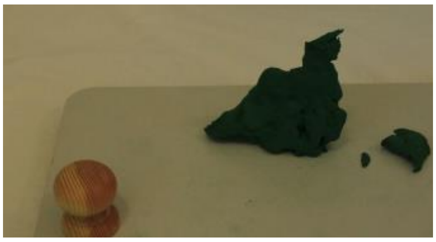
E said “Das ist unser erstes/nächstes Spiel. In dem Becher ist eine Kugel, kannst du sie hören? Da ist ein Sticker an der Kugel. Wenn du die Kugel da rausholen kannst, kannst du den Sticker behalten! Schau mal, da ist schon ein Loch im Deckel. Denkst du, du kannst das größer machen?“

Termite-fish/Tree-hole tool-use

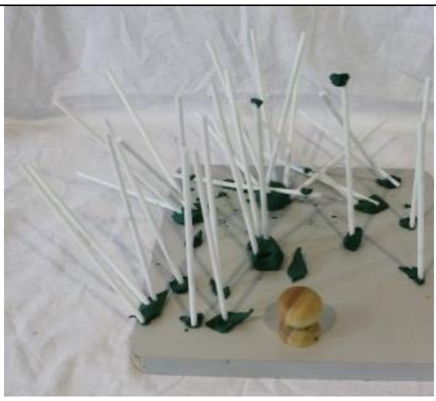
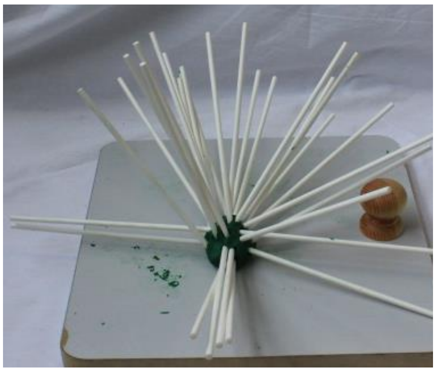
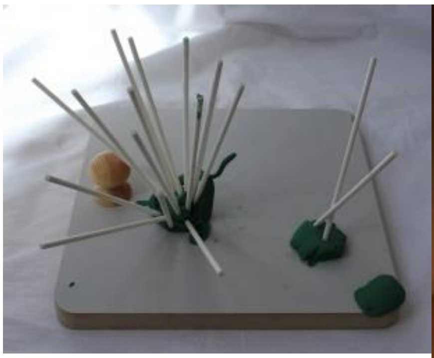
E said “Das ist unser erstes/nächstes Spiel. Hier ist eine Box und da sind drei Sterne. Schau, was ich mache! Eins, zwei, drei! Wenn du die Sterne da rausholen kannst, gewinnst du einen Sticker!“

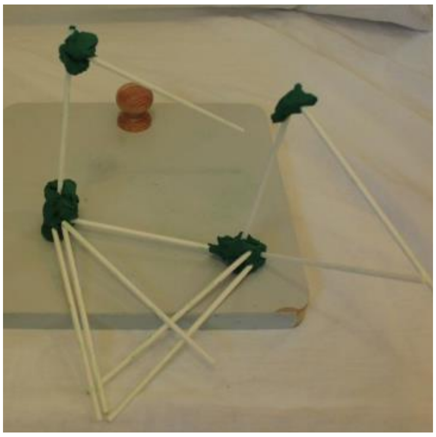
Appendix 3: Towers made in Cumulative culture study 1

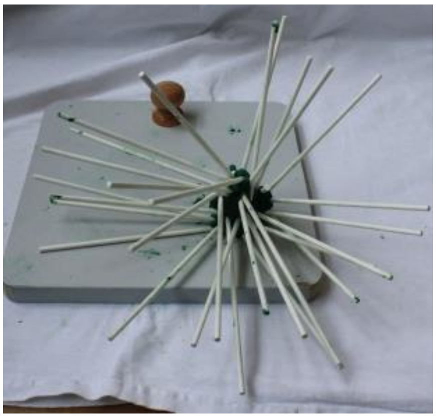
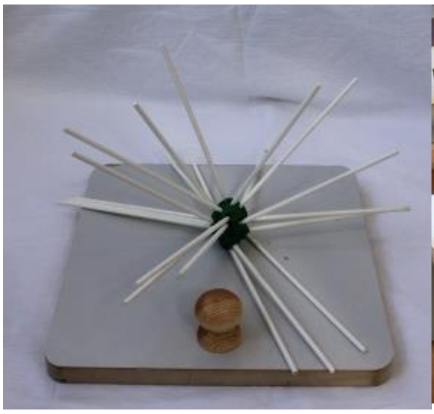
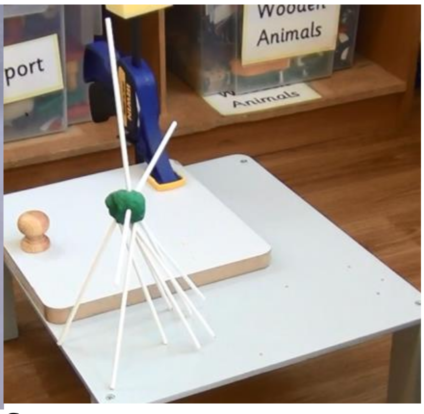
Table A3-1. Towers of children in the baseline condition, ordered by height of final tower.

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if different from final tower) *Picture* *Shape*
16	51	M	 Level-2-tower that never stood on its own 3.5 cm	
15	56	M	 Level-2-tower that never stood on its own 4 cm	
12	62	M	 Bundle 4.5 cm	
13	56	F	 Plasticine tower	

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if different from final tower) *Picture* *Shape*
			6 cm	
6	56	F	 Level-1-Failed 13 cm	 Star
7	54	F	 Star 13.5 cm	

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if different from final tower) *Picture* *Shape*
35	56	F	 Level-1-tower 15.5 cm	
9	67	F	 Hedgehog 16 cm	
3	54	F	 Hedgehog 16 cm	 Hedgehog with plasticine cap on top built

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if different from final tower) *Picture* *Shape*
2	63	F	 Boat 16.5 cm	
14	50	F	 Square 16.5 cm	

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if different from final tower) *Picture* *Shape*
4	58	F	 Hedgehog 16.7 cm	
11	63	M	 Hedgehog 17.3 cm	
1	67	F	 Failed star 20 cm	 Star

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if different from final tower) *Picture* *Shape*
5	57	F	 Star 25 cm	
10	65	M	 Failed level-4-tower (level-2) 27 cm	 Level-3-tower

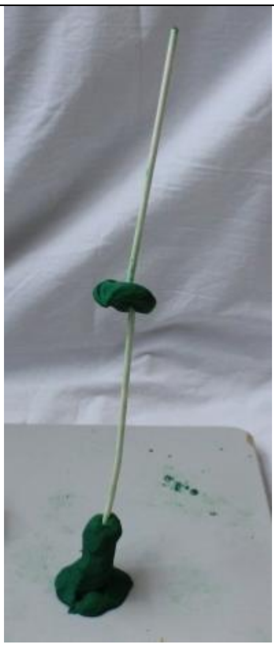
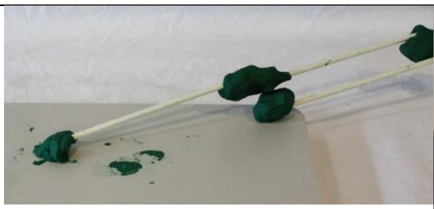
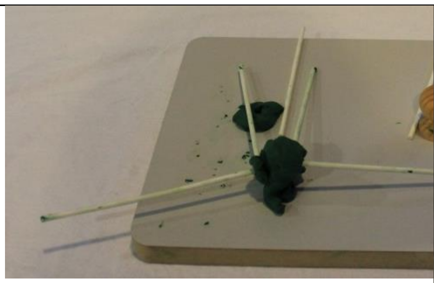
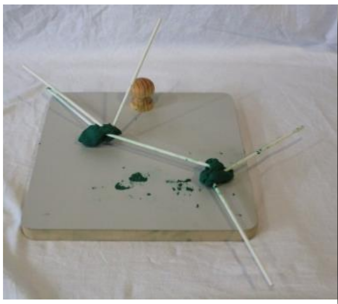
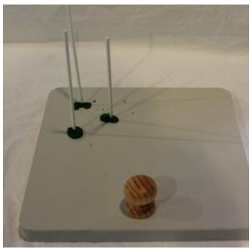
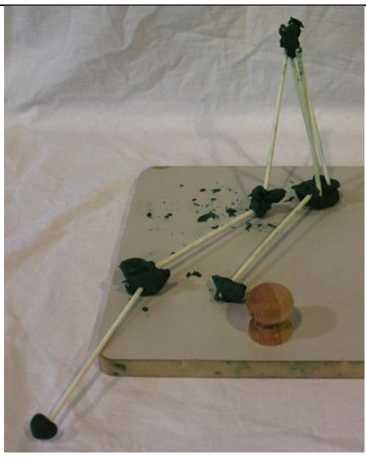
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if different from final tower) *Picture* *Shape*
8	63	M	 Level-2-tower 33.5 cm	

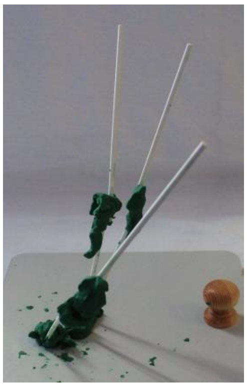
Table A3-2. Towers of children in the action demonstration condition, ordered by height of final tower.

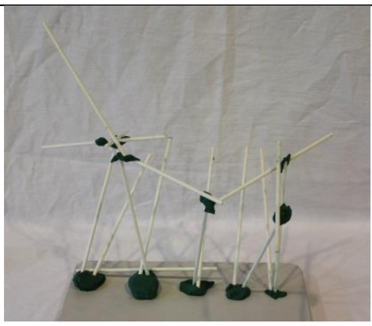
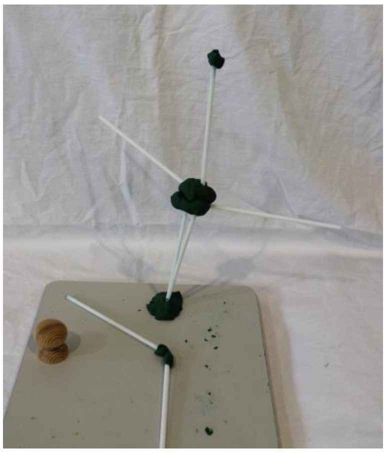
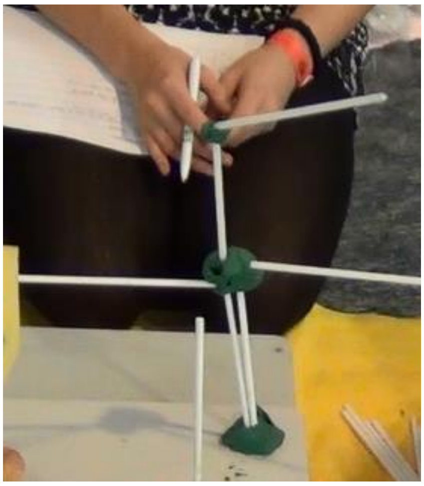
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape*
22	49	F	 Crashed level-3-tower 4 cm	 Level-2-tower

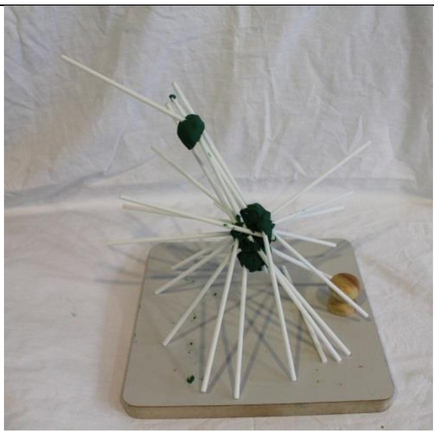
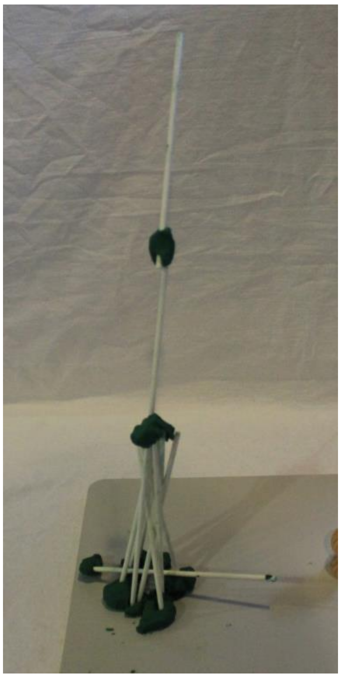
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape*
19	59	M	 Level-1-tower that never stood on its own 4.5 cm	 Level-2-tripod
23	57	M	 Other level-1-construction 11 cm	 Level-1-tower

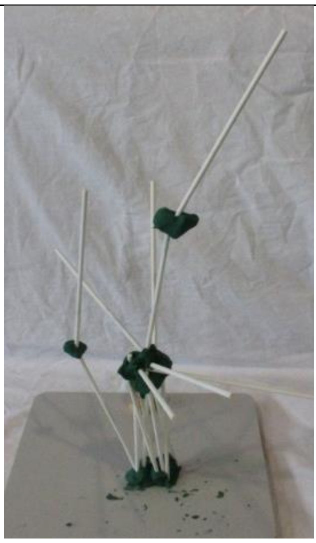
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape*
24	55	F	 Level-1-tower 14 cm	 Level-2-tower
17	54	F	 Level-1-tower 15 cm	 Level-1-tower with plasticine on top

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape*
26	54	M	 Level-4-tower that never stood on its own 16 cm	 Level-2-tripod
34	53	F	 Hedgehog 16 cm	 Level-2-tower

Subject	**Age** in months	Sex	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape*
33	52	M	 Level-1-tower 16.5 cm	 Level-2-tower
21	55	F	 Level-2-tower 27 cm	

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape*
28	62	M	 Level-2-tower 27 cm	
25	61	F	 Level-2-tripod 28.5 cm	
30	61	M	 Level-2-tower 29.5 cm	 Level-3-tower

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape*
29	65	M	 Level-3-star 32 cm	
18	69	M	 Modified tripod 40.5 cm	

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape*
27	58	F	 Modified tripod 42 cm	
32	59	F	 Tripod 43 cm	

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape*
20	55	M	 Tripod 44 cm	

Appendix 4: Establishing the validity of measuring tower height through visual judgement

One of the methodological changes we introduced for cumulative culture study 2 was to measure the height of children's towers continuously throughout the building process, as opposed to only measuring it once at the end of the trial as was done in study 1. We aimed to measure tower height at any time that children added to their construction a new item which increased the tower's height and to take the measurement without disrupting the building process.

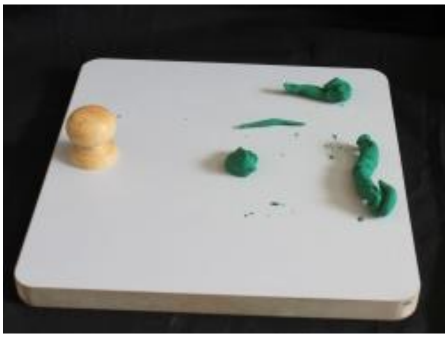
We placed a stationary folding rule at the side of the table, opposite of where E sat. Whenever a child added to the construction an item that increased tower height, E compared the tower to the folding rule in the back to estimate its height. We conducted six trials in a pilot study to determine how precisely E was able to measure tower height via visual judgement. For this, E first built a tower of a random height and shape, after which she measured its height first by visual judgement and then using a folding rule held directly next to the constructions (as was the method for measuring the towers at the end of the trial). The results showed that measurement by visual judgement was a reliable method, with only minimal measurement error (Table A4-1). The general procedure for the study was to take measurements as conservatively as possible. Thus, whenever E was in doubt which of two measurements to take (usually 1 cm apart, e.g., 31.5 or 32.5 cm), she took the higher measurement for the baseline and the lower measurement for the demonstration conditions, so that she would not artificially increase children's performance difference between the baseline and the two experimental conditions (because E was not blind to the hypotheses).

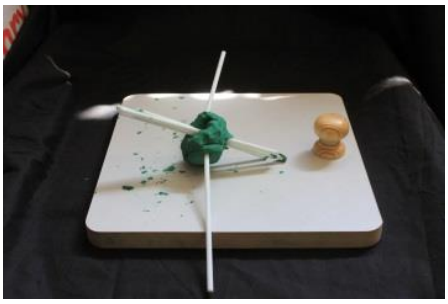
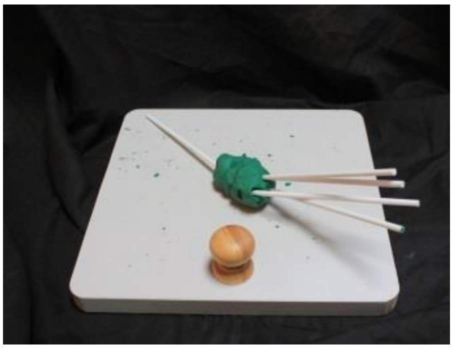
Table A4-1. Pilot data on the accuracy of measuring tower height by visual judgement.

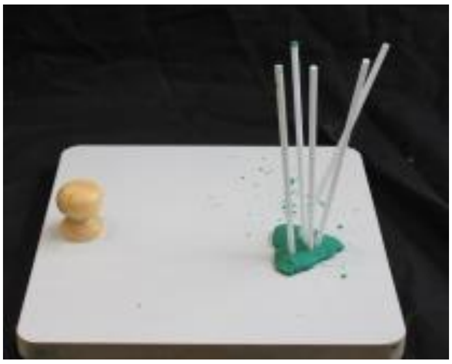
	Estimated height (measured by visual judgement)	**Actual height** (measured with folding rule)	**Estimation error**
Tower 1	19.5 cm	19.5 cm	0.0 cm
Tower 2	16.0 cm	15.5 cm	0.5 cm
Tower 3	41.0 cm	40.0 cm	1.0 cm
Tower 4	31.0 cm	31.0 cm	0.0 cm
Tower 5	57.0 cm	57.5 cm	0.5 cm
Tower 6	7.0 cm	6.5 cm	0.5 cm

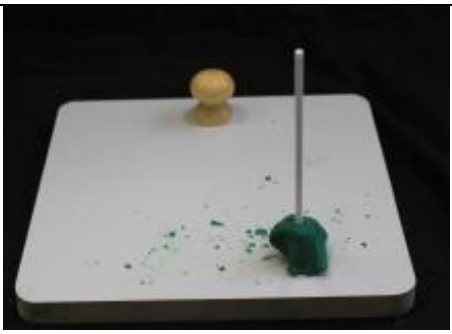
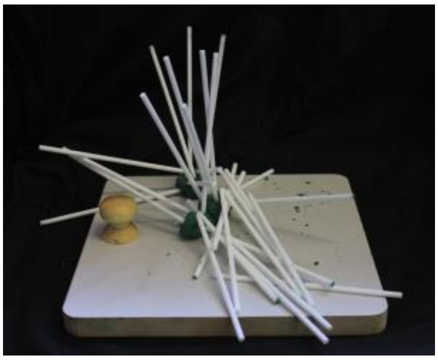
Appendix 5: Towers made in Cumulative culture study 2

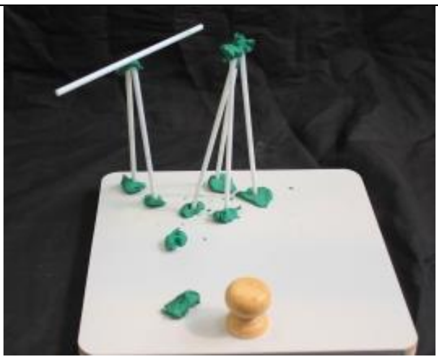
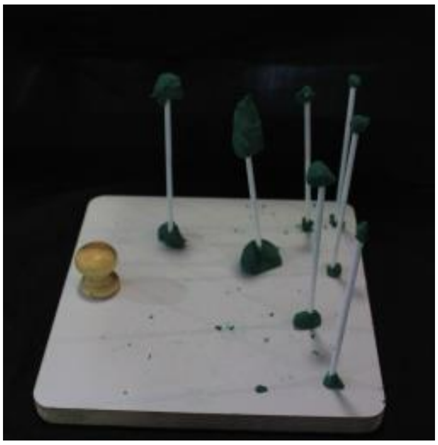
Table A5-1. *Towers of children in the baseline condition, sorted by height of final tower.*

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if different from final tower) *Picture* *Shape* *Height*
93	59	F	 Level-0-construction 0.3 cm	
10	68	F	 Plasticine only 2 cm	
37	50	F	 Level-0-construction 3.4 cm	

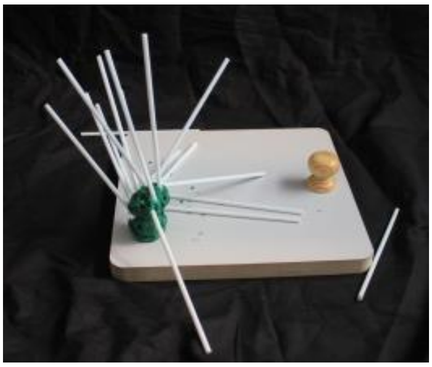
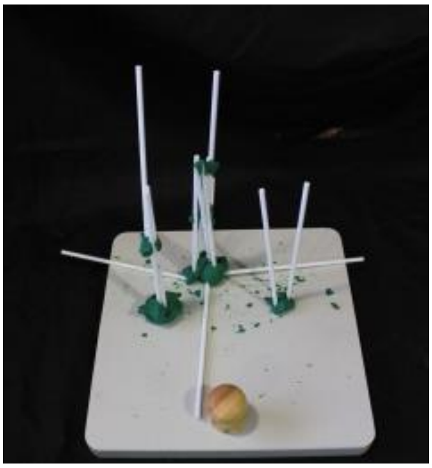
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if different from final tower) *Picture* *Shape* *Height*
91	63	F	 Plasticine tower 8 cm	
45	56	M	 Level-0-construction 7 cm	
55	60	M	 Level-2-tower which never stood on its own 10 cm	No video available Star 17 cm

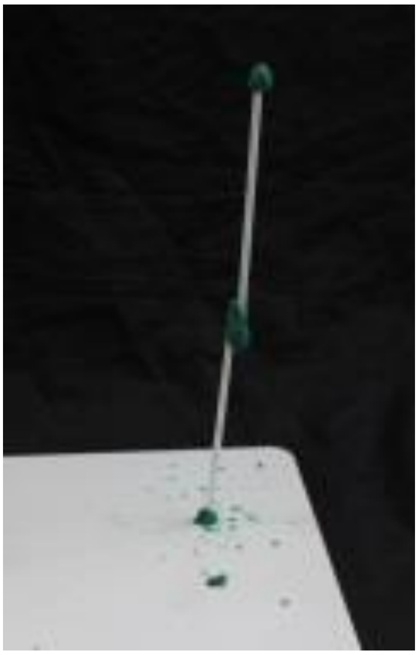
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if different from final tower) *Picture* *Shape* *Height*
75	63	F	 Level-1-tower 15 cm	
26	51	F	 Level-1-tower 15.2 cm	
52	66	M	 Level-1-tower 15.5 cm	

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if different from final tower) *Picture* *Shape* *Height*
32	54	F	 Level-1-tower 15.5 cm	
46	54	F	 Hedgehog 16 cm	
42	55	M	 Hedgehog 16 cm	

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if different from final tower) *Picture* *Shape* *Height*
18	64	M	 Level-1-tower 16.2 cm	
60	64	M	 Level-1-tower 16.5 cm	 Level-3-tower 41 cm
49	63	M	 Level-1-tower	No video available Level-2-tower 28 cm

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if different from final tower) *Picture* *Shape* *Height*
			17 cm	
40	58	F	 Level-1-tower 17 cm	
6	62	F	 Hedgehog 17.8 cm	

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if different from final tower) *Picture* *Shape* *Height*
92	60	F	 Hedgehog 18 cm	
16	65	M	 Hedgehog 19.7 cm	
94	67	F	 Level-2-tower 23 cm	

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if different from final tower) *Picture* *Shape* *Height*
23	61	M	 Level-2-tower 28 cm	
12	57	F	 Level-2-tower 29.8 cm	

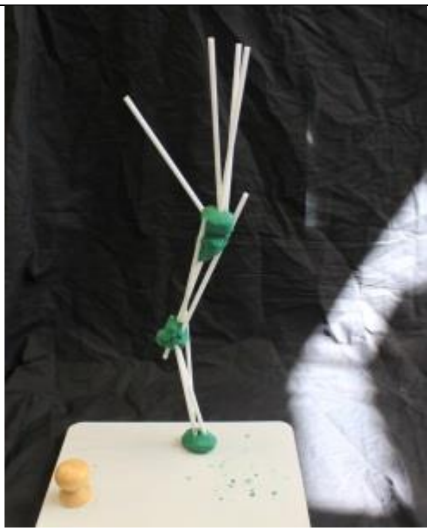
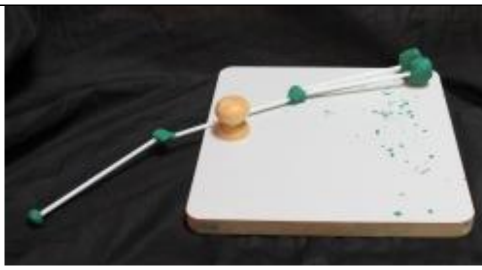
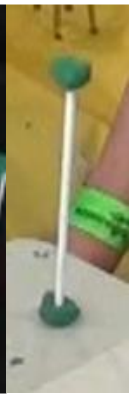
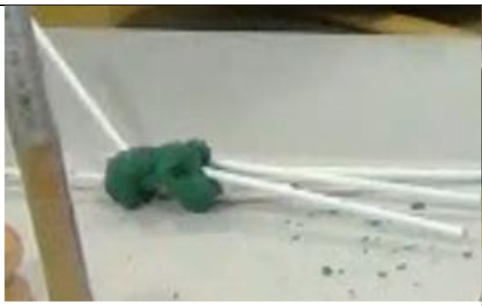
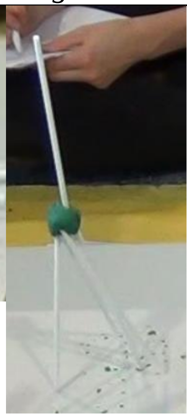
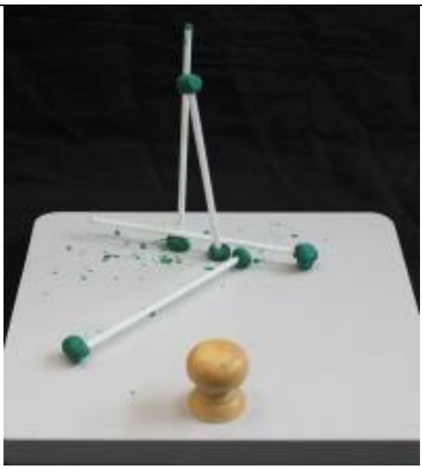
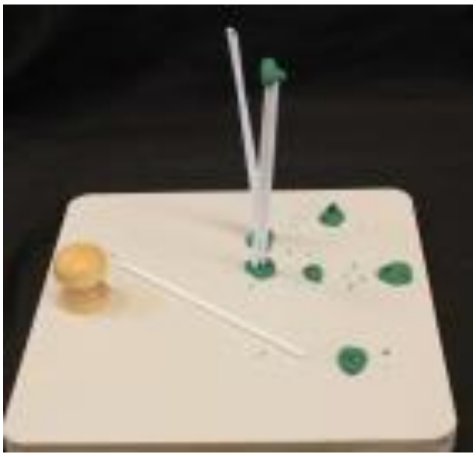
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if different from final tower) *Picture* *Shape* *Height*
20	58	M	 Level-3-tower 41 cm	

Table A5-2. *Towers of children in the full demonstration condition, sorted by height of final tower.*

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
47	60	M	 Crashed level-2-tower 3 cm	No video available Level-2-tower 45 cm
56	63	M	 Plasticine only 3.5 cm	No video available Level-2-tower 29 cm
86	65	M	 Level-0-construction 4 cm	 Level-1-tower 16 cm

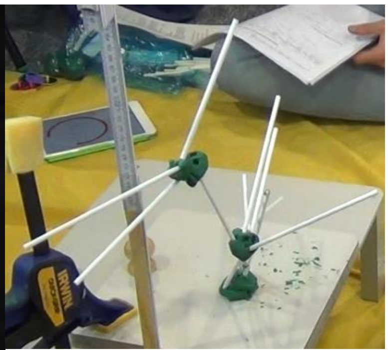
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
66	60	M	 Star-shaped construction which never stood on its own 4.5 cm	 Star 27 cm
57	61	F	No picture available as child cleared table after measuring Level-1-tripod 13 cm	No video available Tripod 41 cm
65	63	M	 Level-1-tower 13.5 cm	 Modified level-2-tripod 32 cm

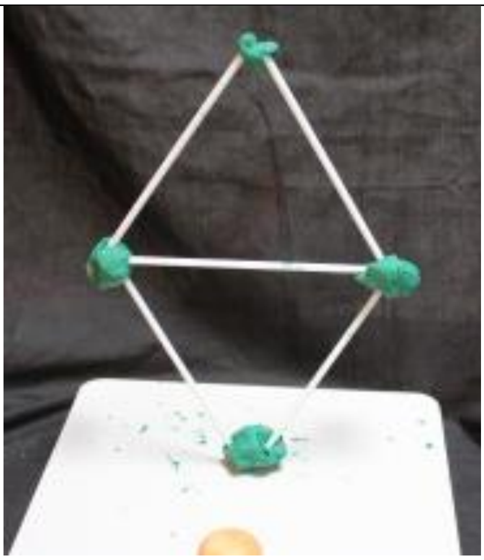
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
80	56	F	 Level-1-tower 15 cm	 Level-2-tower 28.5 cm
43	57	M	 Level-1-tower 15 cm	

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
31	51	F	 Level-1-tower 15 cm	
13	63	M	 Level-1-tower 15.5 cm	

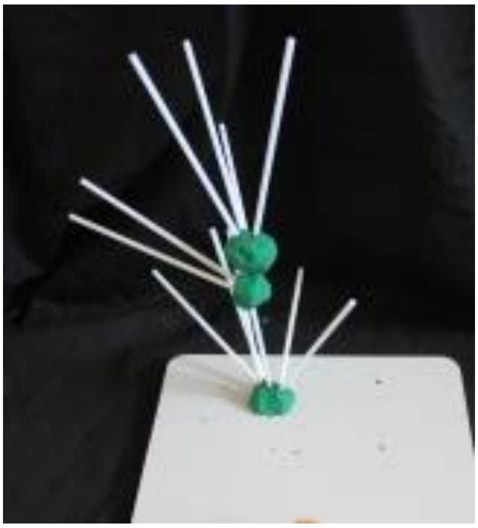
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
53	59	M	 Level-1-tower 15.5 cm	No video available Tripod 43 cm
28	57	F	 Hedgehog 15.8 cm	

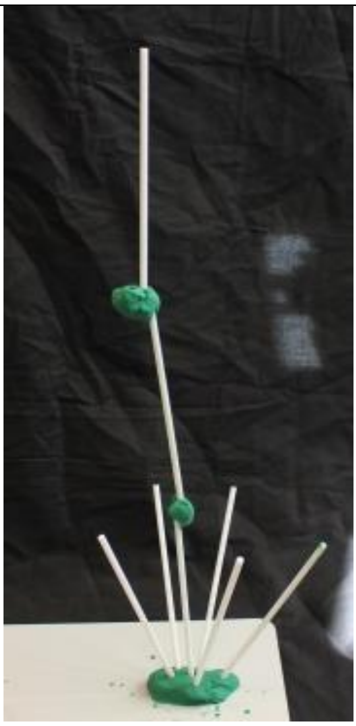
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
38	57	F	 Level-1-tower 15.8 cm	
87	58	M	 Level-2-tower 16 cm	

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
17	60	F	 Hedgehog 16.6 cm	
71	56	M	 Level-2-tower 25 cm	 Level-3-tower 42.5 cm

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
51	58	M	 Diamond 26.5 cm	
41	59	F	 Level-2-tower 28.5 cm	

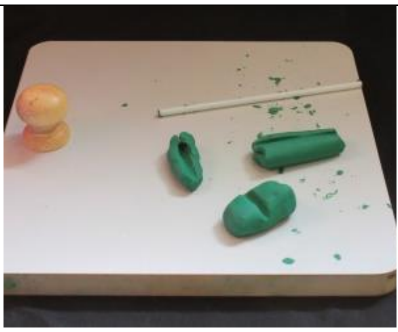
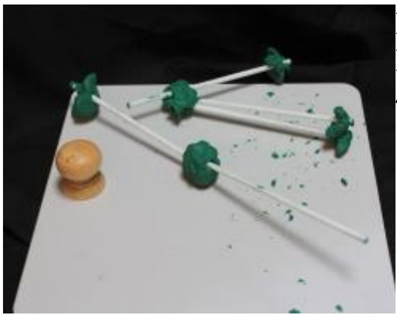
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
62	60	M	 Modified level-2-tripod 29 cm	

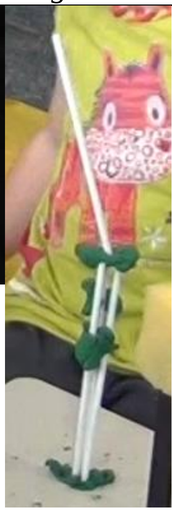
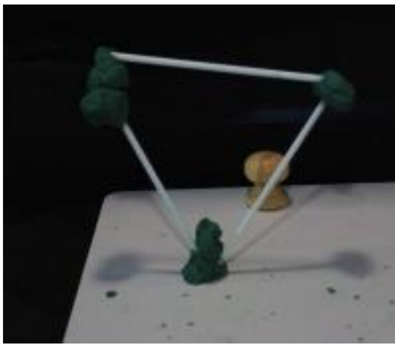
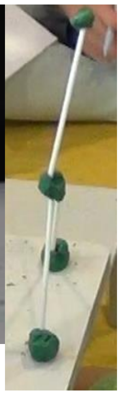
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
24	60	F	 Level-2-tower 30 cm	
7	61	M	 Level-2-tower 30 cm	

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
21	63	F	 Level-3-tower 44 cm	

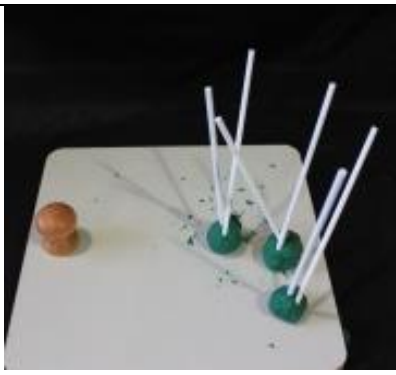
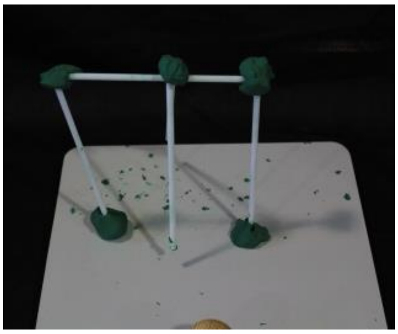
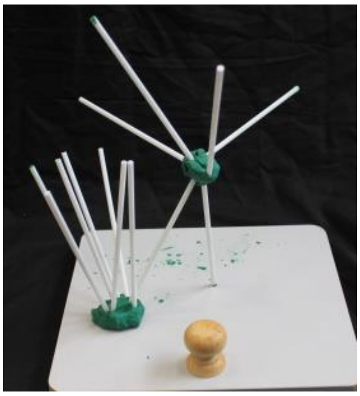
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
64	58	F	 Tripod 44 cm	

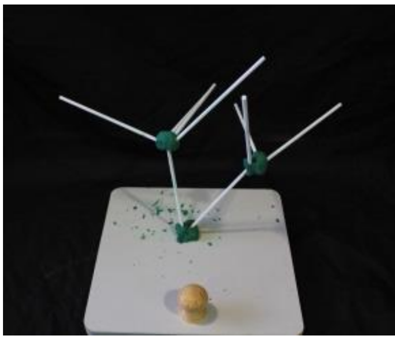
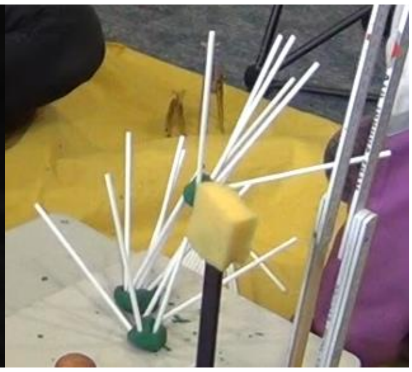
Table A5-3. *Towers of children in the endstate-only demonstration condition, sorted by height of final tower.*

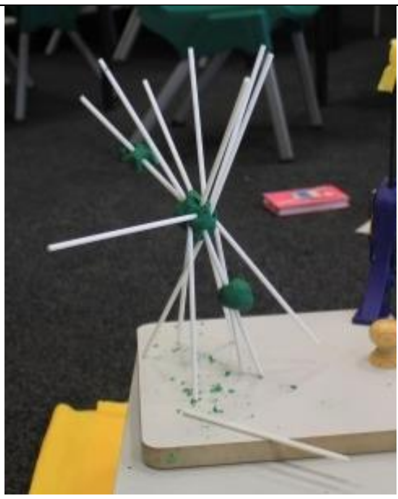
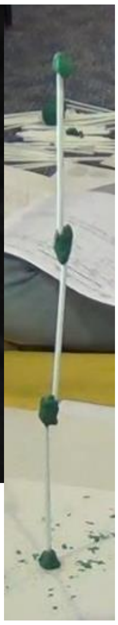
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
44	53	M	 Plasticine tower 2 cm	
54	65	M	 Tripod which never stood on its own 4 cm	No video available Level-3-tower 41 cm
58	64	F	 Crashed tripod 4 cm	No video available Level-2-tower 29 cm

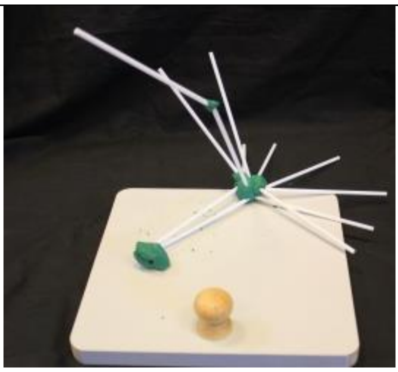
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
84	55	F	 Crashed level-1-tower 6 cm	 Level-2-tower 29 cm
48	64	M	 Level-1-tower 14 cm	
59	57	M	 Level-1-triangle 14.5 cm	 Level-2-tower 28 cm

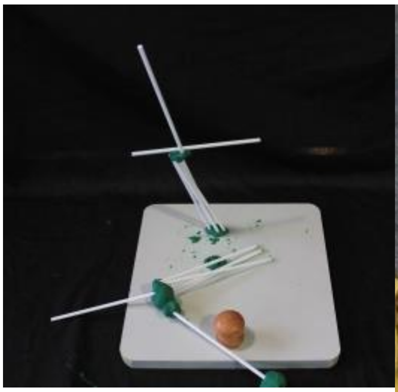
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
73	53	F	 Level-1-cube 15 cm	 Level-2-cube 29 cm
1	61	F	 Level-1-tower 15 cm	
67	57	M	 Level-1-tower 15 cm	

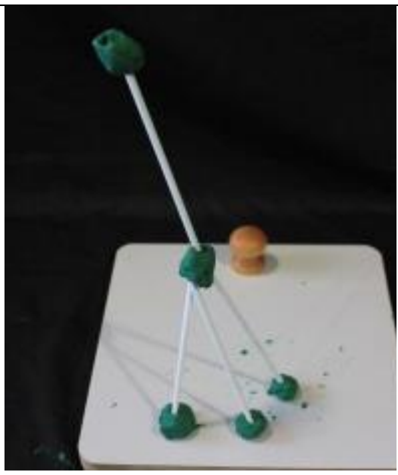
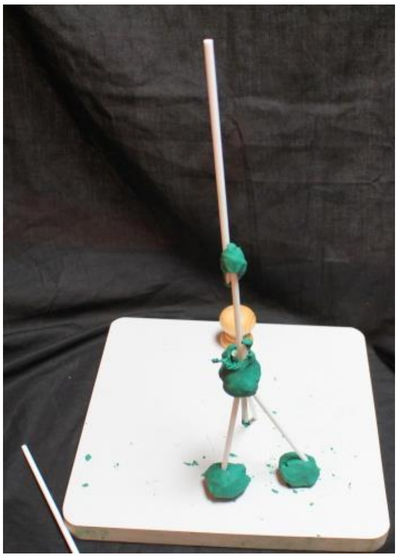
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
90	54	F	 Level-1-tower 15.5 cm	 Level-1-tower 17 cm
63	56	M	 Level-1-square 17 cm	
27	58	M	 Star 21 cm	

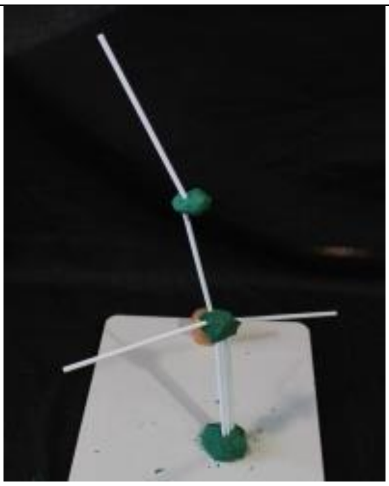
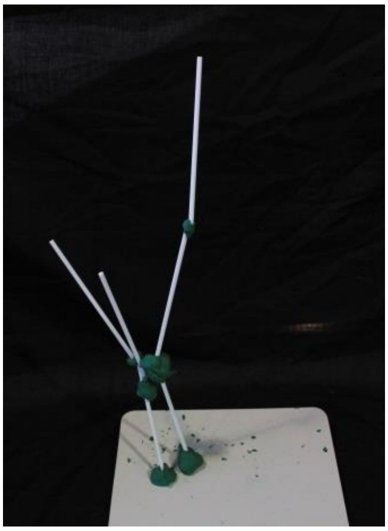
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
69	63	M	 Level-2-tower 23 cm	 Level-2- tower 29 cm
72	64	F	 Level-2-tower 24.5 cm	 Level-2-tower 26 cm
89	63	F	 Level-2-tower 24.7 cm	 Level-2-tower 27.5 cm

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
36	51	F	 Star 28 cm	
74	68	F	 Level-2-tower 28 cm	 Level-3-tower 42 cm

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
14	63	M	 Level-2-tower 28.5 cm	
88	60	F	 Level-2-tower 29 cm	 Level-2-tower 30 cm

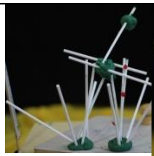
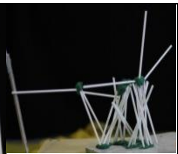
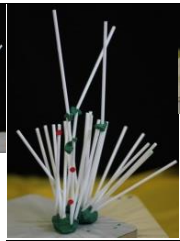
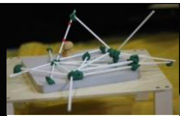
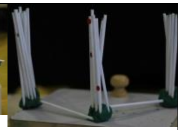
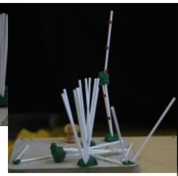
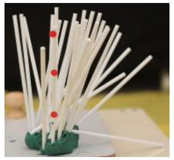
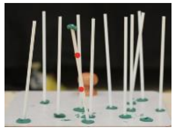
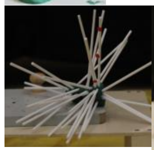
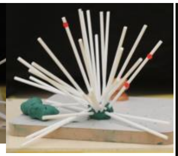
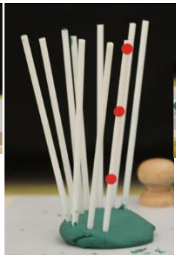
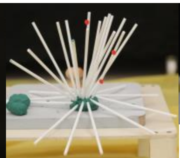
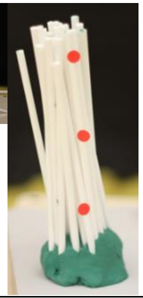
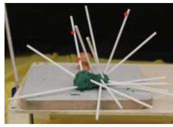
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
81	67	F	 Level-2-tripod 29 cm	
85	64	F	 Level-2-tower 29.5 cm	 Level-3-tower 41 cm

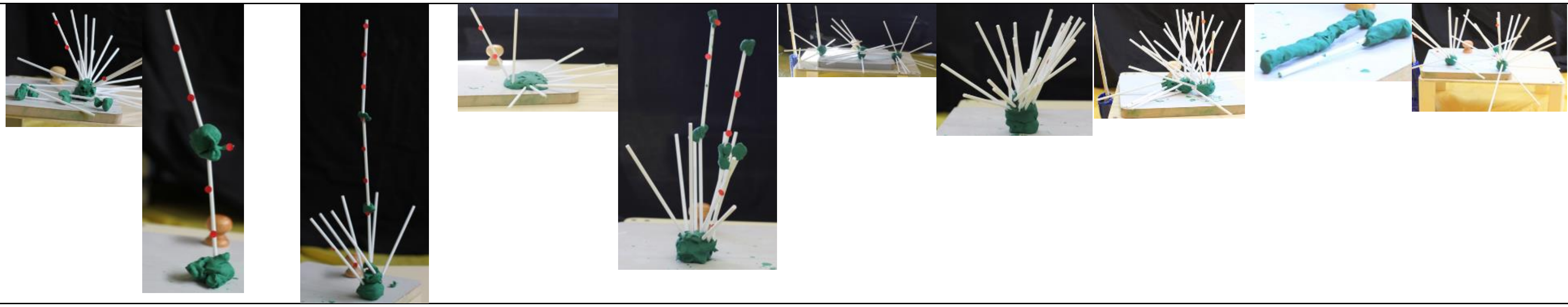
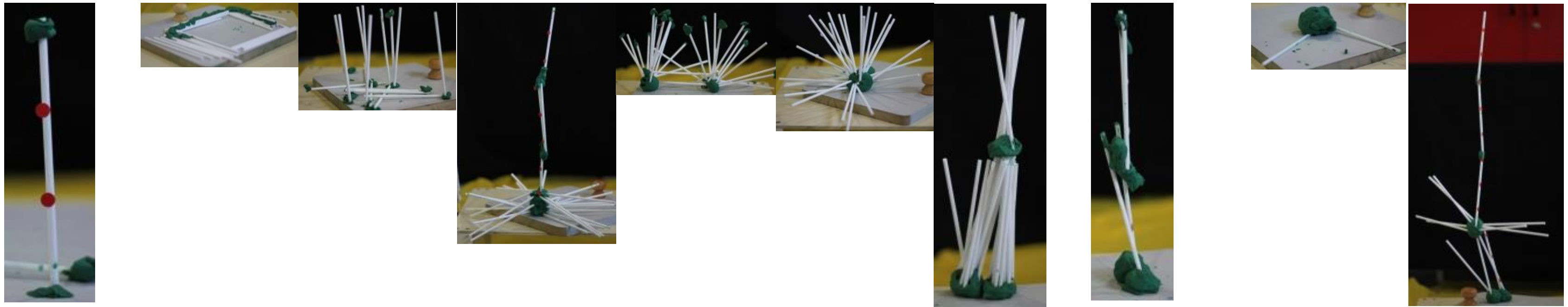
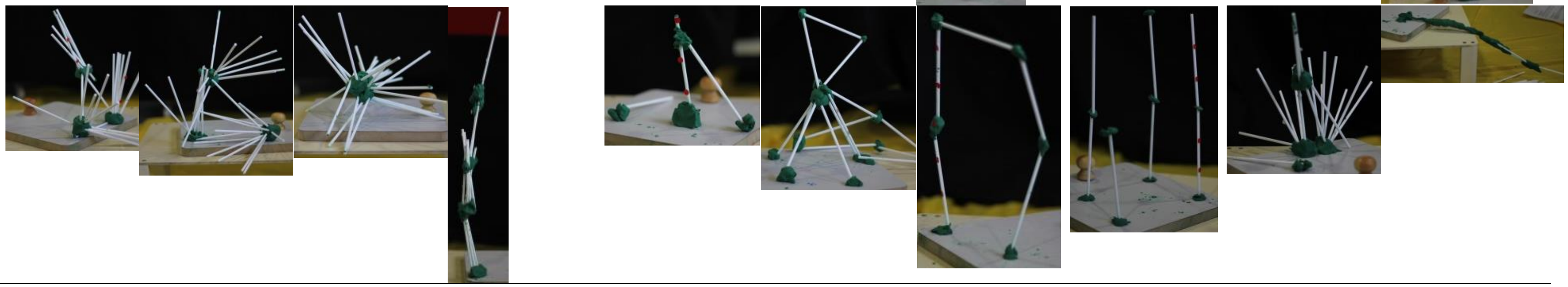
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
83	65	F	 Level-2-tripod 30.5 cm	
50	59	M	 Level-3-tower 33.7 cm	

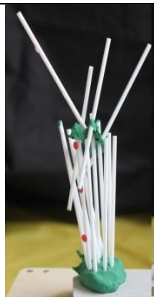
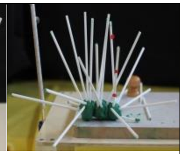
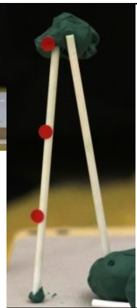
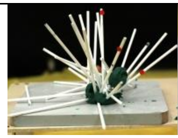
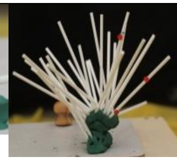
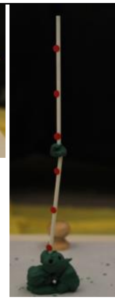
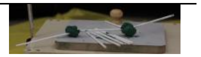
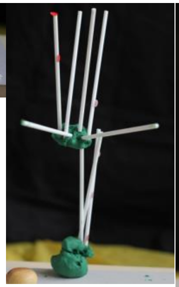
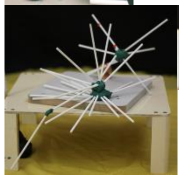
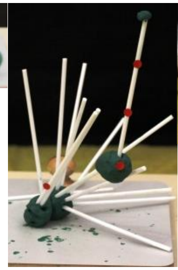
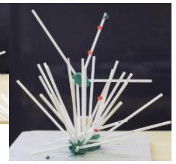
Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
82	60	M	 Level-3-tower 38 cm	
68	63	F	 Level-3-tower 43 cm	

Subject	**Age** in months	**Sex**	**Final tower** *Picture* *Shape* *Height*	**Tower with maximum height** (if final tower was not the one with maximum height) *Picture* *Shape* *Height*
77	57	F	 Tripod 44cm	
29	65	F	 Level-4-tripod 45.5 cm	

Appendix 6: Pictures of towers in the transmission chains

Chain	Position									
Girls	1	2	3	4	5	6	7	8	9	10
1										
2										
3										

Chain	Position									
Girls	1	2	3	4	5	6	7	8	9	10
4										
Boys										
5										
6										

Chain	Position									
Girls	1	2	3	4	5	6	7	8	9	10
7										
8										

LIST OF REFERENCES

Acerbi, A., Jacquet, P. O., & Tennie, C. (2012). Behavioral constraints and the evolution of faithful social learning. *Current Zoology 58*(2), 307-318.

Acerbi, A., & Mesoudi, A. (2015). If we are all cultural Darwinians what's the fuss about? Clarifying recent disagreements in the field of cultural evolution. *Biology and Philosophy, 30*, 481-503.

Acerbi, A., & Tennie, C. (2016). The role of redundant information in cultural transmission and cultural stabilization. *Journal of Comparative Psychology, 130*(1), 62-70.

Alem, S., Perry, C. J., Zhu, X., Loukola, O. J., Ingraham, T., Søvik, E., Chittka, L. (2016). Associative mechanisms allow for social learning and cultural transmission of string pulling in an insect. *PLOS Biology, 14*(10), e1002564.

Alpert, A. (1928). *The solving of problem-situations by preschool children. An analysis.* PhD thesis, Teachers College, Columbia University, New York City.

Ambrose, S. H. (2001). Paleolithic technology and human evolution. *Science, 291*, 1748.

Anderson, J. R., & Henneman, M.-C. (1994). Solutions to a tool-use problem in a pair of *Cebus apella*. *Mammalia*, 58(3), 351–362.

Baayen, R. H. (2008). *Analyzing Linguistic Data. A practical introduction to statistics using R.* New York: Cambridge University Press.

Bandini, E., & Tennie, C. (2017). Spontaneous reoccurrence of "scooping", a wild tool-use behaviour, in naïve chimpanzees. *PeerJ, 5*. e3814.

Bandini, E., Neadle, D., & Tennie, C. (in prep.). A tool-use behaviour, picking, emerges spontaneously in great apes.

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

Barr, R., Dowden, A., & Hayne, H. (1996). Developmental changes in deferred imitation by 6- to 24-month-old infants. *Infant Behavior and Development, 19,* 159-170.

Basalla, G. (1988). *The evolution of technology.* Cambridge: Cambridge University Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-6.

Bates, E., Carlson-Luden, V., & Bretherton, I. (1980). Perceptual aspects of tool using in infancy. *Infant Behavior and Development, 3,* 127-140.

Beck, S. R., Apperly, I. A., Chappell, J., Guthrie, C., & Cutting, N. (2011). Making tools isn't child's play. *Cognition, 119*(2), 301-306.

Beck, S. R., Chappell, J., Apperly, I. A., & Cutting, N. (2012). Tool innovation may be a critical limiting step for the establishment of a rich tool-using culture: a perspective from child development. *Behavioral and Brain Sciences 35*(4), 220-1.

Beck, S. R., Williams, C., Cutting, N., Apperly, I. A., & Chappell, J. (2016). Individual differences in children's innovative problem-solving are not predicted by divergent thinking or executive functions. *Philosophical Transactions of the Royal Society B, 371*, 20150190.

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

Boehm, C. (n.d.). *Variance reduction and the evolution of social control.* Retrieved from http://tuvalu.santafe.edu/files/gems/coevolutionV/boehm.pdf

Boesch, C. (1991). Teaching among wild chimpanzees. *Animal Behaviour, 41*, 530-532.

Boesch, C. (2003). Is culture a golden barrier between human and chimpanzee? *Evolutionary Anthropology, 12*, 82-91.

Boesch, C. (2012). *Wild cultures: A comparison between chimpanzee and human cultures.* New York, USA: Cambridge University Press.

Boesch, C., & Tomasello, M. (1998). Chimpanzee and human cultures. *Current Anthropology, 39*(5), 591-614.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2008). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology and Evolution, 24*(3), 127-135.

Bolwig, N. (1963). Observations on the mental and manipulative abilities of a captive baboon (*Papio doguera*). *Behaviour, 22*(1), 24-40.

Boyd, R., & Richerson, P. J. (1996). Why culture is common, but cultural evolution is rare. *Proceedings of the British Academy, 88*, 73–93.

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 of the United States of America, 108*(2), 10918-10925.

Brown, A. L. (1990). Domain-specific principles affect learning and transfer in children. *Cognitive Science, 14,* 107-133.

Buskell, A. (2017). What are cultural attractors? *Biology & Philosophy,* 1–18.

Byrne, R. W. (1995). *The thinking ape. Evolutionary origins of intelligence.* New York, USA: Oxford University Press.

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

Caldwell, C. A., Cornish, H., & Kandler, A. (2015). Identifying innovation in laboratory studies of cultural evolution: rates of retention and measures of adaptation. *Philosophical Transactions of the Royal Society B, 371*, 20150193.

Caldwell, C. A., & Eve, R. M. (2014). Persistence of contrasting traditions in cultural evolution: Unpredictable payoffs generate slower rates of cultural change. *PLoS ONE 9*, e99708.

Caldwell, C. A., & Millen, A. E. (2008a). Experimental models for testing hypotheses about cumulative cultural evolution. *Evolution and Human Behavior 29,* 165-171.

Caldwell, C. A., & Millen, A. E. (2008b). Studying cumulative cultural evolution in the laboratory. *Philosophical Transactions of the Royal Society B, 363,* 3529–3539.

Caldwell, C. A., & Millen, A. E. (2009). Social learning mechanisms and cumulative cultural evolution: Is imitation necessary? *Psychological Science*, *20*(12), 1478-1483.

Caldwell, C. A., Schillinger, K., Evans, C. L., & Hopper, L. M. (2012). End state copying by humans (*Homo sapiens*): Implications for a comparative perspective on cumulative culture. *Journal of Comparative Psychology, 126,* 161-169.

Call, J. (2013). Three ingredients for becoming a creative tool user. In C. M. Sanz & C. Boesch (Eds.), *Tool use in animals. Cognition and Ecology* (pp. 3-20). New York: Cambridge University Press.

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*, 151-163.

Carpenter, M., & Call, J. (2002). The chemistry of social learning. *Developmental Science, 5,* 22-24.

Carr, K., Kendal, R. L., & Flynn, E. G. (2015). Imitate or innovate? Children's innovation is influenced by the efficacy of observed behavior. *Cognition*, *142*, 322-332.

Carr, K., Kendal, R. L., & Flynn, E. G. (2016). Eureka!: What is innovation, how does it develop, and who does it? *Child Development*, *87*, 1505-1519.

Carvalho, S., Cunha, E., Sousa, C., & Matsuzawa, T. (2008). Chaînes opératoires and resource-exploitation strategies in chimpanzee (*Pan troglodytes*) nut cracking. *Journal of Human Evolution*, *55*, 148-163.

Casler, K., Eshleman, A., Greene, K., & Terziyan, T. (2011). Children's scale errors with tools. *Developmental Psychology, 47*(3), 857– 866.

Casler, K., Kelemen, D. (2005). Young children's rapid learning about artifacts. *Developmental Science, 8*(6), 472-480.

Chappell, J., Cutting, N., Apperly, I. A., & Beck, S. R. (2013). The development of tool manufacture in humans: What helps young children make innovative tools? *Philosophical Transactions of the Royal Society B*, 368, 20120409.

Charbonneau, M. (2015). All innovations are equal, but some more than others: (Re)integrating modification processes to the origins of cumulative culture. *Biological Theory, 10*(4), 322-335.

Chen, Z., Siegler, R. S., & Daehler, M. W. (2000). Across the great divide: Bridging the gap between understanding of toddlers' and older children's thinking. *Monographs of the Society for Research in Child Development, 65*(2), i-vii, 1-96.

Chouinard, M. M. (2007).Children's questions: A mechanism for cognitive development. *Monographs of the Society for Research in Child Development, 72*(1), vii-ix, 1-126.

Claidière, N., & Sperber, D. (2007). The role of attraction in cultural evolution. *Journal of Cognition and Culture, 7*(1), 89–111.

Claidière, N., Scott-Phillips, T. C., Sperber, D. (2014). How Darwinian is cultural evolution? *Philosophical Transactions of the Royal Society B, 369*, 20130368.

Clay, Z., & Tennie, C. (2017). Is overimitation a uniquely human phenomenon? Insights from human children as compared to bonobos. *Child Development*.

Clegg, J. M., & Legare, C. H. (2016). Instrumental and conventional interpretations of behavior are associated with distinct outcomes in early childhood. *Child Development, 87*(2), 527-542.

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In M. H. Johnson & Y. Munakata (Eds.), *Processes of change in brain and cognitive development* (pp. 249-74). Attention and Performance, XXI. Oxford: Oxford University Press.

Connolly, K., & Dalgleish, M. (1989). The emergence of a tool-using skill in infancy. *Developmental Psychology, 25*(6), 894-912.

Coolidge, F. L., & Wynn, T. (2001). Executive functions of the frontal lobes and the evolutionary ascendancy of *Homo sapiens*. *Cambridge Archaeological Journal, 11*(2), 255-260.

Cosmides, L., & Tooby, J. (2001). Unraveling the enigma of human intelligence: Evolutionary psychology and the multimodular mind. In R. J. Sternberg & J. C. Kaufman (Eds.), *The evolution of intelligence* (pp. 145-198). Hillsdale, NJ: Erlbaum.

Custance, D. Whiten, A., & Fredman, T. (1999). Social learning of an artificial fruit task in capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology, 113*(1), 13-23.

Cutting, N., Apperly, I. A., Chappell, J., & Beck, S. R. (2014). The puzzling difficulty of tool innovation: Why can’t children piece their knowledge together? *Journal of Experimental Child Psychology, 125*, 110-117.

Deák, G. O. (2014). Development of adaptive tool-use in early childhood: sensorimotor, social, and conceptual factors. In J. B. Benson (Ed.), *Advances in Child Development and Behavior, 46* (pp. 149-181). Burlington: Academic Press.

Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B., & Laland, K. N. (2012). Identification of the social and cognitive processes underlying human cumulative culture. *Science, 335,* 1114-1118.

Dean, L. G., Vale, G. L., Laland, K. N., Flynn, E., & Kendal, R. L. (2014). Human cumulative culture: A comparative perspective. *Biological Reviews, 89,* 234-301.

Deaner, R. O., van Schaik, C. P., & Johnson, V. (2006). Do some taxa have better domain-general cognition than others? A meta-analysis of nonhuman primate studies. *Evolutionary Psychology, 4,* 149-196.

de Beaune, S. A. (2004). The invention of technology. Prehistory and cognition. *Current Anthropology, 45*(2), 139-162.

Derex, M., & Boyd, R. (2015). The foundations of the human cultural niche. *Nature Communications, 6*(8398).

Derex, M., Godelle, B., & Raymond, M. (2012). Social learners require process information to outperform individual learners. *Evolution, 67,* 688-697.

Enquist, M., Ghirlanda, S., Jarrick, A., & Wachtmeister, C. A. (2008). Why does human culture increase exponentially? *Theoretical Population Biology, 74*(1), 46-55.

Enquist, M., Strimling, P., Eriksson, K., Laland, K., & Sjostrand, J. (2010). One cultural parent makes no culture. *Animal Behaviour, 79,* 1353e1362.

Flynn, E. (2008). Investigating children as cultural magnets: Do young children transmit redundant information along diffusion chains? *Philosophical Transactions of the Royal Society B, 363,* 3541-3551.

Flynn, E., Turner, C., & Giraldeau, L.-A. (2016). Selectivity in social and asocial learning: Investigating the prevalence, effect and development of young children's learning preferences. *Philosophical Transactions of the Royal Society B, 371,* 20150189.

Flynn, E., & Whiten, A. (2008). Cultural transmission of tool use in young children: A diffusion chain study. *Social Development, 17*(3), 699-718).

Flynn, E., & Whiten, A. (2010). Studying children's social learning experimentally "in the wild". *Learning and Behavior, 38*(3), 284-296.

Forstmeier, W., & Schielzeth, H. (2011), Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curve. *Behavioral Ecology & Sociobiology, 65,* 47-55.

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). Thousand Oaks, CA: Sage.

Fragaszy, D. M., & Perry, S. (2003). *The biology of traditions: models and evidence.* Cambridge, UK: Cambridge University Press.

Frye, D., Clark, A., Watt, D., & Watkins, C. (1986). Children's construction of horizontals, verticals, and diagonals: an operational explanation of the "Oblique Effect". *Developmental Psychology, 22*(2), 213-217.

Galef, B. G. (2009). Culture in animals? In K. N. Laland & B. G. Galef (Eds.), *The question of animal culture* (p. 222-246). Cambridge, MA: Harvard University Press.

Galef, B. G. (2012). Social learning and traditions in animals: Evidence, definitions, and relationship to human culture. *WIREs Cognitive Science, 3,* 581-592.

Gardiner, A. K., Bjorklund, D. F., Greif, M. L., & Gray, S. K. (2012). Choosing and using tools: prior experience and task difficulty influence preschoolers‘ tool-use strategies. *Cognitive Development*, *27*, 240-254.

Gentner, D., Levine, S. C., Ping, R., Isaia, A., Dhillon, S., Bradley, C., & Honke, G. (2015). Rapid learning in a children's museum via analogical comparison. *Cognitive Science, 40*(1), 224-240.

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature, 415*, 755.

Gergely, G., & Csibra, G. (2006). Sylvia’s recipe: The role of imitation and pedagogy in the transmission of cultural knowledge. In: N. J. Enfield & S. C. Levenson (Eds.), *Roots of Human Sociality: Culture, Cognition, and Human Interaction* (pp. 229-255). Oxford: Berg Publishers.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3½-7 years old on a Stroop-like day-night test. *Cognition, 53*(2), 129-153.

Gwilliams, A., Reindl, E., & Tennie, C. (2017, May). *Did a puzzlebox identify the social and cognitive processes underlying human cumulative culture?* Poster session presented at Culture Conference, Birmingham, UK.

Haidle, M. N. (2010). Working-memory capacity and the evolution of modern cognitive potential. Implications from animal and early human tool use. *Current Anthropology, 51*(1), S149-S166.

Haidle, M. N., Bolus, M., Collard, M., Conard, N. J., Garofoli, D., Lombard, M., … Whiten, A. (2015). The Nature of culture: An eight-grade model for the evolution and expansion of cultural capacities in hominins and other animals. *Journal of Anthropological Sciences*, *93*, 43-70.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, *21*, 803-865.

Harmand, S., Lewis, J. E., Feibel, C. S., Lepre, C. J., Prat, S., Lenoble, A., … Roche, H. (2015). 3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya. *Nature, 521*, 310-315.

Harris, J. W. K. (1983). Cultural beginnings: Plio-Pleistocene archaeological occurrences from the Afar, Ethiopia. *The African Archaeological Review, 1*, 3-31.

Harris, P. L. (2012). *Trusting what you're told: How children learn from others.* Cambridge, MA: Harvard University Press.

Hashimoto, C., Isaji, M., Koops, K., & Furuichi, T. (2015). First records of tool-set use for ant-dipping by Eastern chimpanzees (*Pan troglodytes schweinfurthii*) in the Kalinzu Forest Reserve, Uganda. *Primates. 56*(4), 301-305.

Henrich, J. (2004). Demography and cultural evolution: How adaptive cultural processes can produce maladaptive losses: the Tasmanian case. *American Antiquity, 69*(2), 197-214.

Henrich, J. (2015). *The secret of our success. How culture is driving human evolution, domesticating our species, and making us smarter*. New Jersey: Princeton University Press.

Henrich, J., & Boyd, R. (2002). On modeling cultural evolution: Why replicators are not necessary for cultural evolution. *Journal of Cognition and Culture, 2*(2), 87-112.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61-135.

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

Heyes, C. M. (1993). Imitation, culture and cognition. *Animal Behaviour, 46*, 999-1010.

Heyes, C. (2012a). What’s social about social learning? *Journal of Comparative Psychology, 126*(2), 193-202.

Heyes, C. (2012b). Grist and mills: On the cultural origins of cultural learning. *Philosophical Transactions of the Royal Society B, 367*, 2181-2191.

Heyes, C. (2013). What can imitation do for cooperation? In K. Sterelny, R. Joyce, B. Calcott, & B. Fraser (Eds.), *Cooperation and its evolution* (pp. 313-332). Cambridge: MIT Press.

Heyes, C. (2016a). Imitation: Not in our genes. *Current Biology 26*, R408–R431.

Heyes, C. (2016b). Born pupils? Natural pedagogy and cultural pedagogy. *Perspectives on Psychological Science, 11*(2), 280-295.

Hill, K. (2010). Experimental studies of animal social learning in the wild: Trying to untangle the mystery of human culture. *Learning & Behavior, 38*(3), 319-328.

Hill, K., Barton, M., & Hurtado, M. (2009). The emergence of human uniqueness: Characters underlying behavioral modernity. *Evolutionary Anthropology, 18*(5), 187–200.

Hill, K. R., Walker, R. S., Božičević, M., Eder, J., Headland, T. Hewlett, B., … Wood, B. (2011). Co-residence patterns in hunter-gatherer societies show unique human social structure. *Science, 331*, 1286-1289.

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.

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,* 486-493.

Hoppitt, W., & Laland, K. N. (2013). *Social learning: An introduction to mechanisms, methods, and models.* Princeton, US: Princeton University Press.

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.

Hothorn, T., Bretz, F., &Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*(3), 346-363.

Huffman, M. A., Hirata, S. (2004). An experimental study of leaf swallowing in captive chimpanzees: Insights into the origin of a self-medicative behavior and the role of social learning. *Primates, 45*(2), 113-118.

Hunt, G. R., Gray, R. D., & Taylor, A. H. (2013). Why is tool use rare in animals? In C. M. Sanz & C. Boesch (Eds.), *Tool use in animals. Cognition and Ecology* (pp. 89-118). New York: Cambridge University Press.

Kahrs, B. A., Jung, W. P., & Lockman, J. J. (2012). What is the role of infant banging in the development of tool use? *Experimental Brain Research, 218*, 315–320.

Kahrs, B. A., Jung, W. P., & Lockman, J. J. (2013). Motor origins of tool use. *Child Development, 84*(3), 810–816,

Kahrs, B. A., & Lockman, J. J. (2014). Tool using. *Child Development Perspectives, 8*(4), 231–236.

Keen, R. (2011). The development of problem solving in young children: A critical cognitive skill. *Annual Review of Psychology, 62*, 1-21.

Keen, R., Lee, M.-H., & Adolph, K. (2014). Planning an action: a developmental progression in tool use, *Ecological Psychology*, *26*(1-2), 98-108,

Kempe, M., & Mesoudi, A. (2014). An experimental demonstration of the effect of group size on cultural accumulation. *Evolution and Human Behavior*, *35*(4), 285-290.

Kenward, B. (2012). Over-imitating preschoolers believe unnecessary actions are normative and enforce their performance by a third party. *Journal of Experimental Child Psychology, 112*(2), 195-207.

Keupp, S., Behne, T., & Rakoczy, H. (2013). Why do children overimitate? Normativity is crucial. *Journal of Experimental Child Psychology, 116*(2), 392-406.

Kitahara-Frisch, J. (1993). The origin of secondary tools. In A. Berthelet & J. Chavaillon (Eds.), *The use of tools by human and non-human primates* (pp. 239-246). New York, USA: Oxford Science Publications.

Köhler, W. (1925). *The mentality of apes* (2nd ed.). USA: Liveright.

Kline, M. A., & Boyd, R. (2010). Population size predicts technological complexity in Oceania. *Proceedings of the Royal Society B, 277,* 2559–2564.

Laland, K. N., & Hoppitt, W. (2003). Do animals have culture? *Evolutionary Anthropology, 12*(3), 150-159.

Laland, K. N., & O’Brien, M. J. (2012). Cultural niche construction: An introduction. *Biological Theory, 6*(3), 191-202.

Laland, K. N., Odling-Smee, J., & Feldman, M. W. (2000). Niche construction, biological evolution, and cultural change. *Behavioral and Brain Sciences, 23,* 131-175.

Langergraber, K. E., Boesch, C., Inoue, E., Inoue-Murayama, M., Mitani, J. C., Nishida, T., …Vigilant, L. (2011). Genetic and 'cultural' similarity in wild chimpanzees. *Proceedings of the Royal Society B, 278*, 408-416.

Legare, C. H., & Harris, P. L. (2016). The ontogeny of cultural learning. *Child Development, 87*(3), 633-642.

Legare, C. H., & Nielsen, M. (2015). Imitation and innovation: The dual engines of cultural learning. *Trends in Cognitive Sciences 19*(11), 688-699.

Legare, C. H., Wen, N, J., Herrmann, P. A., & Whitehouse, H. (2015). Imitative flexibility and the development of cultural learning. *Cognition*, *142*, 351-361.

Lehner, S. R., Burkart, J. M., & van Schaik, C. P. (2010). An evaluation of the geographic method for recognizing innovations in nature, using zoo orangutans. *Primates 51*(2), 101-118.

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, 367*, 2171-2180.

Lockman, J. J. (2000). A perception-action perspective on tool use development. *Child Development*, *71*(1), 137-144.

Lockman, J. J. (2005). Tool use from a perception–action perspective: developmental and evolutionary considerations. In V. Roux & B. Bril (Eds.), *Stone knapping: The necessary conditions for a uniquely hominin behavior* (pp. 319–330). Cambridge, UK: McDonald Institute for Archaeological Research.

Logan, C. J., Breen, A. J., Taylor, A. H., Gray, R. D., & Hoppitt, W. J. E. (2015). How New Caledonian crows solve novel foraging problems and what it means for cumulative culture. *Learning & Behavior 44*(1), 18-28.

Lombard, M., & Haidle, M. N. (2012). Thinking a bow-and-arrow set: Cognitive implications of middle stone age bow and stone-tipped arrow technology. *Cambridge Archaeological Journal, 22*(2), 237-264.

Luncz, L. V., Mundry, R., & Boesch, C. (2012). Evidence for cultural differences between neighboring chimpanzee communities. *Current Biology, 22*(10), 922–926.

Luria, A. R., & Vygotsky, L. (1930). *Ape, primitive man, and child: Essays in the history of behaviour.* London: Harvester Wheatsheaf.

Lyons, D. E., Young, A. G., & Keil, F. C. (2007).The hidden structure of overimitation. *Proceedings of the National Academy of Sciences USA 104*(50), 19751-19756.

Mannu, M., & Ottoni, E. B. (2009). The enhanced tool-kit of two groups of wild bearded capuchin monkeys in the Caatinga: Tool making, associative tool use, and secondary tools. *American Journal of Primatology, 71*(3), 242-251.

Marsh, L., Ropar, D., & Hamilton, A. (2014). The social modulation of imitation fidelity in school-age children. *PLoS ONE, 9,* e86127.

Martin-Ordas, G., Schumacher, L., & Call, J. (2012). Sequential tool use in great apes. *PLoS ONE, 7*(12), 1-15.

Matheson, E. (1931). A study of problem solving behavior in pre-school children. *Child Development*, *2*(4), 242-262.

Matsuzawa, T. (1991a). Nesting cups and metatools in chimpanzees. *Behavioral and Brain Sciences, 14*(4), 570-571.

Matsuzawa, T. (1994). Field experiments on use of stone tools in the wild. In R. W. Wrangham., W. C. McGrew, F. B. M. de Waal, & P. G. Heltne (Eds.), *Chimpanzee cultures* (pp. 351-370). Cambridge, USA: Harvard University Press.

McCarty, M. E., Clifton, R. K., & Collard, R. R. (2001). The beginnings of tool use by infants and toddlers. *Infancy, 2*(2), 233-256.

McGrew, W. C. (1987). Tools to get food: The subsistants of Tasmanian Aborigines and Tanzanian chimpanzees compared. *Journal of Anthropological Research, 43*(3), 247-258.

McGuigan, N. (2013). The influence of model status on the tendency of young children to over-imitate. *Journal of Experimental Child Psychology, 116*(4), 962-969.

McGuigan, N., Burdett, E., Burgess, V., Dean, L., Lucas, A., Vale, G. & Whiten, A. (in press). Innovation and social transmission in experimental micro-societies: exploring the scope of cumulative culture in young children. *Philosophical Transactions of the Royal Society B.*

McGuigan, N., & Graham, M. (2010). Cultural transmission of irrelevant tool actions in diffusion chains of 3- and 5-year-old children. *European Journal of Developmental Psychology, 7*(5), 561-577.

McGuigan, N., & Robertson, S. (2015). The influence of peers on the tendency of 3- and 4-year-old children to over-imitate. *Journal of Experimental Child Psychology, 136,* 42-54.

McGuigan, N., & Whiten, A. (2009). Emulation and "overemulation" in the social learning of causally opaque versus causally transparent tool use by 23- and 30-month-olds. *Journal of Experimental Child Psychology, 104*(4), 367-381.

Menzel, C., Fowler, A., Tennie, C., & Call, J. (2013). Leaf surface roughness elicits leaf swallowing behavior in captive chimpanzees (*Pan troglodytes*) and bonobos (*P. paniscus*), but not in gorillas (*Gorilla gorilla*) or orangutans (*Pongo abelii*). *International Journal of Primatology, 34*(3), 533-553.

Mesoudi, A. (2007). Using the methods of experimental social psychology to study cultural evolution. *Journal of Social, Evolutionary & Cultural Psychology, 1*(2), 35-58.

Mesoudi, A. (2011). *Cultural evolution: How Darwinian theory can explain human culture and synthesize the social sciences.* Chicago, IL: University of Chicago Press.

Mesoudi, A., & Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society B, 363,* 3489-3501.

Migliano, A. B., Page, A. E., Gómez-Gardeñes, J., Salali, G. D., Viguier, S., Dyble, M., … Vinicius, L. (2017). Characterization of hunter-gatherer networks and implications for cumulative culture. *Nature Human Behaviour, 1*(0043), 1-6.

Möbius, Y., Boesch, C., Koops, K., Matsuzawa, T., & Humle, T. (2008). Cultural differences in army ant predation by West African chimpanzees? A comparative study of microecological variables. *Animal Behaviour, 76*(1), 37–45.

Moore, R. (2013). Imitation and conventional communication. *Biology & Philosophy, 28*(3), 481-500.

Morgan, T. J. H., Uomini, N. T., Rendell, L. E., Chouniard-Thuly, L., Street, S. E., Lewis, H., … Laland, K. N. (2015). Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature Communications, 6,* 6029.

Morin, O. (2016a). *How traditions live and die.* New York, US: Oxford University Press.

Morin, O. (2016b). Reasons to be fussy about cultural evolution. *Biology & Philosophy, 31*(3), 447e458.

Mulcahy, N. J., Call, J. (2006). Apes save tools for future use. *Science, 312,* 1038-1040.

Mulcahy, N. J., Call, J., & Dunbar, R. I. M. (2005). Gorillas (*Gorilla gorilla*) and orangutans (*Pongo pygmaeus*) encode relevant problem features in a tool-using task. *Journal of Comparative Psychology, 119*(1), 23-32.

Muthukrishna, M., & Henrich, J. (2016). Innovation in the collective brain. *Philosophical Transactions of the Royal Society B, 371* (1690).

Muthukrishna, M., Shulman, B. W., Vasilescu, V., & Henrich, J. (2014) Sociality influences cultural complexity. *Proceedings of the Royal Society B, 281*, 20132511.

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

NASA (2016, October 11). *Journey to Mars Overview.* Retrieved from https://www.nasa.gov/content/journey-to-mars-overview

Nielsen, M. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychol*ogy*, 42*(3)*,* 555-565.

Nielsen, M., & Blank, C. (2011). Imitation in young children: When who gets copied is more important than what gets copied. *Developmental Psychology, 47*(4), 1050-1053.

Nielsen, M., Kapitány, R., & Elkins, R. (2015). The perpetuation of ritualistic actions as revealed by young children's transmission of normative behavior. *Evolution and Human Behavior, 36*(3)*,* 191–198.

Nielsen, M., Tomaselli, K., Mushin, I., & Whiten, A. (2014). Exploring tool innovation: A comparison of Western and Bushman children. *Journal of Experimental Child Psychology, 126*, 384-394.

O'Brien, M. J., Lyman, R. L., Mesoudi, A., & VanPool, T. L. (2010). Cultural traits as units of analysis. *Philosophical Transactions of the Royal Society B, 365*, 3707-3806.

Oostenbroeck, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., … Slaughter, V. (2016). Comprehensive longitudinal study challenges the existence of neonatal imitation in humans. *Current Biology, 26*(10), 1334–1338.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science 349*, 6251.

Osiurak, F., De Oliveira, E., Navarro, J., Lesourd, M., Claidiere, N., & Reynaud, E. (2016). Physical intelligence does matter to cumulative technological culture. *Journal of Experimental Psychology 145*(8), 941-948.

Over, H., & Carpenter, M. (2012). Putting the social into social learning: Explaining both selectivity and fidelity in children's copying behavior. *Journal of Comparative Psychology, 126*(2), 182–192.

Page, E. B. (1963). Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of the American Statistical Association, 58*(301), 216–230.

Piaget, J. (1952). *The origins of intelligence in children.* New York: International University Press.

Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences, 107*(2), 8993-8999.

Petroski, H. (1997). *The evolution of useful things.* New York: Vintage Books.

Powell, A., Shennan, S. J., & Thomas, M. G. (2009). Late Pleistocene demography and the appearance of modern human behavior. *Science 324*, 1298–1301.

R Core Team (2013/2015). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: https://www.R-project.org

Rat-Fischer, L., O'Regan, J. K., & Fagard, J. (2012). The emergence of tool use during the second year of life. *Journal of Experimental Child Psychology, 113*(3), 440-446.

Ray, E., & Heyes, C. (2011). Imitation in infancy: The wealth of the stimulus. *Developmental Science 14*(1), 92–105.

Read, D. W. (2008). Working-memory: A cognitive limit to non-human primate recursive thinking prior to hominid evolution. *Evolutionary Psychology, 6*(4), 676-714.

Reindl, E., Bandini, E., & Tennie, C. (in press). The zone of latent solutions and its relation to the classics: Vygotsky and Köhler. In L. D. Di Paolo & F. D. Vincenzo (Eds.) *Social cognition in non-human primates and early Homo*. Berlin: Springer.

Rendell, L., & Whitehead, H. (2001). Culture in whales and dolphins. *Behavioral and Brain Sciences, 24*(2), 309-382.

Richerson, P. J., & Boyd, R. (2005). *Not by genes alone. How culture transformed human evolution.* Chicago, IL: Chicago University Press.

Robbins, M. M., Ando, C., Fawcett, K. A., Grueter, C. C., Hedwig, D., Iwata, Y., … Yamagiwa, J. (2016). Behavioral variation in gorillas: Evidence of potential cultural traits. *PLoS ONE, 11*(9): e0160483.

Ruiz A. M., & Santos, L. R. (2013). Understanding differences in the way human and non-human primates represent tools: The role of teleological-intentional information. In C. M. Sanz & C. Boesch (Eds.), *Tool use in animals. Cognition and Ecology* (pp. 119-133). New York: Cambridge University Press.

Rybanska, V., McKay, R., Jong, J., & Whitehouse, H. (2017). Rituals improve children's ability to delay gratification. *Child Development.* doi:10.1111/cdev.12762

Sasaki, T., & Biro, D. (2017). Cumulative culture can emerge from collective intelligence in animal groups. *Nature Communications*, *8,* 15049.

Schick, K., & Toth, N. (2000). Origin and development of tool-making behavior in Africa and Asia. *Human Evolution*, *15*(1-2), 121-128.

Schillinger, K., Mesoudi, A., & Lycett, S. J. (2015). The impact of imitative versus emulative learning mechanisms on artefactual variation: Implications for the evolution of material culture. *Evolution and Human Behavior 36*(6), 446-455.

Schöning, C., Humle, T., Möbius, Y., & McGrew, W. C. (2008). The nature of culture: Technological variation in chimpanzee predation on army ants revisited. *Journal of Human Evolution 55*(1), 48-59.

Scott-Phillips, T. C. (2017). A (simple) experimental demonstration that cultural evolution is not replicative, but reconstructive – and an explanation of why this difference matters. *Journal of Cognition and Culture 17*(1-2), 1-11.

Shattuck, R. (1980). *The forbidden experiment: The story of the wild boy of Aveyron.* Pennsylvania: Kodansha International.

Sheridan, K. M., Konopasky, A. W., Kirkwood, S., & Defeyter, M. A. (2016). The effects of environment and ownership on children's innovation of tools and tool material selection. *Philosophical Transactions of the Royal Society B, 371,* 20150191.

Shumaker R. W., Walkup, K.R., & Beck, B.B. (2011). *Animal tool behavior. The use and manufacture of tools by animals*. Baltimore: Johns Hopkins University Press.

Slater, P. J. B. (1986). The cultural transmission of bird song. *Trends in Ecology and Evolution, 1*, 94–97.

Smith, C. M. (2016). An adaptive paradigm for human space settlement. *Acta Astronautica, 119,* 207-217.

Somogyi, E., Ara, C, Gianni, E., Rat-Fischer, L., Fattor, P., O'Regan, J. K., & Fagard, J. (2015). The roles of observation and manipulation in learning to use a tool. *Cognitive Development, 35,* 186-200.

Spence, K. W. (1937). Experimental studies of learning and the mental processes in infra-human primates. *Psychological Bulletin, 34*(10), 806–50.

Sperber, D. (1996). *Explaining culture: A naturalistic approach.* Oxford: Blackwell.

Stadler, K. (2015). *Page's test is not a trend test.* Retrieved from: https://github.com/kevinstadler/notes/raw/master/page-test.pdf

Stadler, K. (2017). cultevo: Tools, measures and statistical tests for cultural evolution. R package. Retrieved from https://kevinstadler.github.io/cultevo/

Sterelny, K. (2012). *The evolved apprentice: How evolution made humans unique*. Cambridge, USA: MIT Press.

Sterelny, K. (2017). Cultural evolution in California and Paris. *Studies in History and Philosophy of Biological and Biomedical Sciences, 62,* 42-50.

Stout, D., & Khreisheh, N. (2015). Skill learning and human brain evolution: An experimental approach. *Cambridge Archaeological Journal, 25*(4), 867-875.

Stout, D., Toth, N., & Schick, K. (2000). Stone tool-making and brain activation: Position emission tomography (pet) studies. *Journal of Archaeological Science, 27*(12), 1215–1223.

Taylor, A. H., Hunt, G. R., Holzhaider, J. C., & Gray, R. D. (2007). Spontaneous metatool use by New Caledonian crows. *Current Biology, 17*(17), 1504-1507.

Tecwyn, E. C., Thorpe, S. K. S., & Chappell, J. (2013). A novel test of planning ability: Great apes can plan step-by-step but not in advance of action. *Behavioural Processes, 100,* 174-184.

Tecwyn, E. C., Thorpe, S. K. S., & Chappell, J. (2014). Development of planning in 4- to 10-year-old children: reducing inhibitory demands does not improve performance. *Journal of Experimental Child Psychology, 125,* 85-101.

Tennie, C., Caldwell, C., & Dean, L. (2017). Cumulative culture. In H. Callan (Ed.), *International Encyclopedia of Anthropology*. Oxford: Wiley-Blackwell.

Tennie, C., Call., & Tomasello, M. (2006). Push or pull: Imitation vs. emulation in great apes and human children. *Ethology, 112*, 1159–1169.

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: On the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B, 364,* 2405-2415.

Tennie, C., Call, J., & Tomasello, M. (2012). Untrained chimpanzees (*Pan troglodytes schweinfurthii*) fail to imitate novel actions. *PLoS ONE, 7* (8), 1-19.

Tennie, C., & Hedwig, D. (2009). How latent solution experiments can help to study differences between human culture and primate traditions. In E. Potocki & J. Krasinski (Eds.), *Primatology: Theories, Methods and Research* (pp. 95-112). New York, US: Nova Publishers.

Tennie, C., Hedwig, D., Call, J., & Tomasello, M. (2008). An experimental study of nettle feeding in captive gorillas. *American Journal of Primatology, 70*(6), 584-593.

Tennie, C., & Over, H. (2012). Cultural intelligence is key to explaining human tool use. *Behavioral and Brain Sciences, 35*(4), 242-243.

Tennie, C., Walter, V., Gampe, A., Carpenter, M., & Tomasello, M. (2014). Limitations to the cultural ratchet effect in young children. *Journal of Experimental Child Psychology, 126*, 152-160.

Thompson, D. E., & Russell, J. (2004). The ghost condition: Imitation versus emulation in young children's observational learning. *Developmental Psychol*ogy, *40*(5), 882-889.

Thorpe, W. (1956). *Learning and instinct in animals.* London: Methuen.

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

Tomasello, M. (1999b). The human adaptation for culture. *Annual Review of Anthropology, 28*, 509-529.

Tomasello, M. (2009). The question of chimpanzee culture, plus postscript (Chimpanzee culture, 2009). In K. N. Laland & B. G. Galef (Eds.), *The question of animal culture* (pp. 198–221). Cambridge, MA: Harvard University Press.

Tomasello, M. (2011). Human culture in evolutionary perspective. In M. Gelfand (Ed.), *Advances in Culture and Psychology* (pp. 5-51). New York, US: Oxford University Press.

Tomasello, M. (2016). The ontogeny of cultural learning. *Current Opinion in Psychology, 8,* 1–4.

Tomasello, M., & Call, J. (2008). Assessing the validity of ape-human comparisons: A reply to Boesch (2007). *Journal of Comparative Psychology, 122*(4), 449-452.

Tomasello, M., & Herrmann, E. (2010). Ape and human cognition: What's the difference? *Current Directions in Psychological Science, 19*(1), 3-8.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences, 16,* 495-552.

Tomasello, M., & Moll, H. (2010). The gap is social: Human shared intentionality and culture. In P. Kappeler & J. Silk (Eds.), *Mind the gap: Tracing the origins of human universals* (pp. 331-349). Berlin: Springer.

Toth, N., & Schick, K. (2009). The Oldowan: The tool making of early hominins and chimpanzees compared. *Annual Review of Anthropology, 38*, 289-305.

Toth, N., Schick, K. D., Savage-Rumbaugh, E. S., Sevcik, R. A., & Rumbaugh, D. M. (1993). Pan the tool-maker: Investigations into the stone tool-making and tool-using capabilities of a bonobo (*Pan paniscus*). *Journal of Archaeological Science, 20*(1), 81-91.

Uzgiris, I. C. (1981). Two functions of imitation during infancy. *International Journal of Behavioral Development, 4*(1), 1-12.

Vaesen, K. (2012). The cognitive bases of human tool use. *Behavioral and Brain Sciences, 35*(4), 203-262.

van Leeuwen, E. J. C., Cronin, K. A., Haun, D. B. M., Mundry, R., & Bodamer, M. D. (2012). Neighbouring chimpanzee communities show different preferences in social grooming behaviour. *Proceedings of the Royal Society B, 279,* 4362–4367.

van Schaik, C. P. (2003). Local traditions in orangutans and chimpanzees: Social learning and social tolerance. In D. M. Fragaszy & S. Perry (Eds.), *The Biology of Traditions: Models and Evidence* (pp. 297-328). Cambridge: Cambridge University Press.

van Schaik, C. P., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C. D., Singleton, I., … Merrill, M. (2003). Orangutan cultures and the evolution of material culture. *Science, 299*, 102-105.

van Schaik, C. P., Ancrenaz, M., Djojoasmoro, R., Knott, C. D., Morrogh-Bernard, H. C., Odom, K., … van Noordwijk, M. A. (2009). Orangutan cultures revisited. In S. A. Wich, S. S. Utami Atmoko, T. Mitra Setia, & C. P. van Schaik (Eds.), *Orangutans: geographic variation in behavioral ecology and conservation.* (pp. 299-309). New York, US: Oxford University Press.

van Schaik, C. P., & Burkart, J. M. (2011). Social learning and evolution: The cultural intelligence hypothesis. *Philosophical Transactions of the Royal Society B, 366,* 1008-1016.

van Schaik, C. P., & Pradhan, G. R. (2003). A model for tool-use traditions in primates: Implications for the coevolution of culture and cognition. *Journal of Human Evolution 44*(6), 645–64.

Völter, C. J., & Call, J. (2014). Younger apes and human children plan their moves in a maze

Task. *Cognition, 130*, 186-203.

Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Want, S. C., & Harris, P. L. (2001). Learning from other people's mistakes: Causal understanding in learning to use a tool. *Child Development, 72*(2), 431-443.

Want, S. C. & Harris, P. L. (2002). How do children ape? Applying concepts from the study of non-human primates to the developmental study of "imitation" in children. *Developmental Science, 5*(1), 1-41.

Wasielewski, H. (2014). Imitation is necessary for cumulative cultural evolution in an unfamiliar, opaque task. *Human Nature 25*(1), 161-179.

Watson-Jones, R. E., Legare, C. H., Whitehouse, H., & Clegg, J. M. (2014). Task-specific effects of ostracism on imitative fidelity in early childhood. *Evolution and Human Behavior, 35*(3), 204-210.

Westergaard, G. C., Greene, J. A., Menuhin-Hauser, C., Suomi, S. J. (1996). The use of naturally-occurring copper and iron tools by monkeys: Possible implications for the emergence of metal-tool technology in hominids. *Human Evolution, 11*(1), 17-25.

Westergaard, G. C., Lundquist, A. L., Kuhn, H. E., & Suomi, S. J. (1997). Ant-gathering with tools by captive tufted capuchins (*Cebus apella*). *International Journal of Primatology, 18*(1), 95-103.

Westergaard, G. C., & Suomi, S. J. (1994). Stone-tool bone-surface modification by monkeys. *Current Anthropology, 35*(4), 468-470.

Whalley, C. L., Cutting, N., & Beck, S. R. (2017). The effect of prior experience on children's tool innovation. *Journal of Experimental Child Psychology, 161*, 81-94.

Whiten, A. (2005). The second inheritance system of chimpanzees and humans. *Nature 437,* 52-55.

Whiten, A., Custance, D. M., Gomez, J.-C., Teixidor, P., & Bard, K. A. (1996). Imitative learning of artificial fruit processing in children (*Homo sapiens*) and chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 110*(1), 3-14.

Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., … Boesch, C. (1999). Cultures in chimpanzees. *Nature, 399,* 682-685.

Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., … Boesch, C. (2001). Charting cultural variation in chimpanzees. *Behaviour, 138*(11), 1481-1516.

Whiten, A., Horner, V., Litchfield, C. A., Marshall-Pescini, S. (2004). How do apes ape? *Learning and Behavior, 32*(1), 36-52.

Whiten, A., McGuigan, N., Marshall-Pescini, S., & Hopper, L. M. (2009). Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philosophical Transactions of the Royal Society B, 364,* 2417-2428.

Whiten, A., & van Schaik, C. P. (2007). The evolution of animal „cultures“ and social intelligence. *Philosophical Transactions of the Royal Society B, 362*, 603-620.

Wilks, M., Kapitány, R., & Nielsen, M. (2016). Preschool children's learning proclivities: When the ritual stance trumps the instrumental stance. *British Journal of Developmental Psychology, 34*(3), 402-14.

Willatts, P. (1984). The stage-IV infant's solution of problems requiring the use of supports. *Infant Behavior and Development, 7,* 125-134.

Wimpenny, J. H., Weir, A. A. S., Clayton, L., Rutz, C., & Kacelnik, A. (2009). Cognitive processes associated with sequential tool use in New Caledonian crows. *PLoS ONE, 4(*8), 1-16.

Wimsatt, W. C. (1999). Genes, memes and cultural heredity. *Biology & Philosophy, 14*(2), 279-310.

Wood, L. A., Kendal, R. L., & Flynn, E. G. (2013). Whom do children copy? Model-based biases in social learning. *Developmental Review 33*(4), 341–356.

Wrangham, R.W. (2001). Out of the Pan, into the fire: How our ancestors' evolution depended on what they ate. In F. B. M. de Waal (Ed.), *Tree of Origin*. (pp. 119-143). Cambridge: Harvard University Press.

Wrangham, R. W. (2009). *Catching fire: How cooking made us human.* New York: Basic Books.

Zink, K. D., & Lieberman, D. E. (2016). Impact of meat and Lower Palaeolithic food processing techniques on chewing in humans. *Nature, 531,* 500–503.

Zwirner, E., & Thornton, A. (2015). Cognitive requirements of cumulative culture: Teaching is useful but not essential. *Scientific Reports 5***,** 16781.