Social learning and creativity in children in informal learning environments

MURSIC, ZARJA

Social learning and creativity in children in informal learning environments

Zarja Muršič

A thesis presented for the degree of Doctor of Philosophy

Anthropology Department
Durham University
2018

Table of contents

List of tables

List of figures

List of abbreviations

The Interactive Research Pod – IRP

Life Science Centre – LSC

Monte Carlo Markov Chain – MCMC

Glossary of commonly used terms

Creativity – something that is novel

Social learning – defined broadly learning from copying others, including imitation and emulation

Structures – builds

Coder – person who counted the behaviours either from the videos or live

Rater – adults recruited either through the University or Prolific that rated pairs of images based on their similarity

Rating – an item on a Likert scale from 1 to 7 describing the similarity of the structures children build or the 'correspondence' of the structures to what children said they built

Declaration

I confirm that no part of the material present in this thesis has previously been submitted for a degree in this or any other institution. If material has been generated through joint work, this has been indicated where appropriate. All other sources have been referenced, and quotations suitably indicated.

Acknowledgements

Rachel Kendal and Jeremy Kendal, thank you for enabling and supporting my exploration of the world of children through the lenses of innovation and social learning.

Andy Lloyd, thank you for opening the doors to the Life Science Centre and for all the support with data collection. I am very grateful also to all the members of staff at the LSC for offering support and help.

The team that worked behind designing the Interactive Research Pod, Rachel Kendal, Andy Lloyd, Hannah Rudman, Claire Bailey-Cross, Bethan Ross and Jeremy Kendal, thank you for fruitful meetings that led to an excellent exhibit without which this thesis would not exist.

Guy Lavender-Forsyth, thank you for your assistance with data collection and for the invaluable discussions during the quiet times at the LCS.

Marie-Claire, thank you for being around.

Ada, Freja and Lajka, thank you for making sure I was walked daily during the last year of the work on this thesis.

Durham friends, you know who you are, thanks for the company along the way.

Family and friends, thank you for supporting me.

Public Scholarship, Development, Disability and Maintenance Fund of the Republic of Slovenia, thank you for financially supporting my studies with the Ad futura Scholarship for Studies Abroad.

Abstract

Social learning, together with innovation, form the pillars of human culture. The vast majority of research regarding innovation and social learning uses artificially created tasks (e. g. puzzle boxes) with clear goals. Studies with children are performed in nurseries, psychology laboratories and in separate rooms in science centres. This enables studies to have high experimental/internal validity. However, it is not known whether these findings also explain behaviours outside of controlled environments.

In this thesis I explored how social learning and creativity could be studied in the context of an informal learning environment (Life Science Centre, Newcastle) using an open-ended task, representing a context of increased ecological validity. In Chapter 3 I explored how direct instructions, scaffolding (open questions) and no instructions impacted children's exploratory behaviour and their creativity when building with shapeshifting wooden blocks that constituted an existing exhibit in the science centre. In Chapter 4 I used an exhibit, the Interactive Research Pod (IRP), which was developed through cooperation between Durham University academics and Life Science Centre practitioners to study social learning and creativity "in the wild" whereby no experimenter is present and instructions, cameras and ethical assent is automated. I studied children, who were using building blocks, in social (transparent partitions), asocial (opaque partitions, building at or around the same time) and asocial control (opaque partitions, different day). In Chapter 5 I used the IRP to enable children to freely interact whilst building and investigated social learning and cooperation as well as the originality of the final structures. In all three studies I used a newly developed web application to evaluate the creativity of the structures children built. I recruited

adult raters to acquire a relatively objective measure of the subjective value of the originality of the wooden structures. I used Bayesian statistical methods to analyse the data.

Overall children built diverse structures and were not strongly impacted by the conditions (direct instruction, scaffolding, no instruction, social, asocial learning and cooperation) they were in. The findings of the thesis complement existing data regarding social learning and creativity in children in more controlled environments and demonstrate the utility of conducting such studies in ecologically valid contexts, despite the inherent issues of internal validity.

1 General Introduction

Children express diverse explorative behaviours (Bonawitz et al. 2011; Bonawitz et al. 2012; Rusher, Cross, and Ware 1995; Van Schijndel and Raijmakers 2016; Shneidman et al. 2016), explore the world as 'little scientists' through experimentation and hypothesis testing (Gopnik 2012), 'little anthropologists' through observing, participating and learning from others (Legare and Harris 2016) and they are very good imitators (Meltzoff 1985; Nielsen, Cucchiaro, and Mohamedally 2012), but generally do not innovate and create useful products (Beck et al. 2011; Cutting, Apperly, and Beck 2011; but see Reindl et al. 2016). Most studies of cultural learning and innovation in children investigated children's behaviours in laboratory or formal educational contexts (Carr, Kendal, and Flynn 2015; Dean et al. 2014; Tennie et al. 2014). The literature on children's cultural learning and creativity in, arguably more common informal learning contexts, is scarce. In this thesis I investigated children's social learning and creativity in a series of naturalistic experiments.

In the first part of this chapter I present definitions of creativity and connect it to the innovation literature. In the second part I consider children as social and cultural learners. In the last part I stress the importance of conducting experiments in diverse environmental and social contexts.

1.1 Creativity and Innovation

Cultural innovations and novel cultural traditions stem from creative processes (Carr, Kendal, and Flynn 2016; Fogarty, Creanza, and Feldman 2015). Generally creativity is understood as something novel (Carruthers 2002). Novelty is sometimes presented also as originality (Carr, Kendal, and Flynn 2016). Usefulness is commonly included in the definition in addition to novelty (Stein 1953; Sternberg 2006). Creativity can be defined and studied on a historical or individual level (Boden 2004). Historical creativity introduces something that is new to humankind (Boden, 2004). It can be understood in a similar way as innovation where novel behaviours are acquired (learned) by others and spread through the population (Carr, Kendal, and Flynn 2016). Creativity on the individual level comprises of processes that are novel and useful to the individual and can also be described as psychological creativity (Boden, 2004).

The sociocultural definition of creativity focuses on its usefulness and describes creativity as “the generation of a product that is judged to be novel and also appropriate, useful, or valuable by a suitably knowledgeable social group” (Sawyer 2012, p. 8) or in other words “ideas, behaviours, or products that are both novel and valuable” (Piccuito and Carruthers 2014, p. 199). It is important to note that, unlike innovation, creativity is not necessarily useful in the functional, technological sense but it can be valuable (e. g. work of art), appropriate (e. g. novel scientific work of a graduate student in their thesis) or efficient (e. g. novel solution to a problem that already has solution) (Stein 1953; Ivcevic 2009; Runco and Jaeger 2012; Abraham 2019). Furthermore, the usefulness might come only after the novel and creative solution emerges and is evaluated based on its usefulness (Diedrich et al. 2015). In addition to historical and individual level creativity, some researchers also divide creativity into big-C and little-c creativity (Kaufman and Beghetto 2009; Stein, 1987 as cited in Sawyer, 2012). Big-C refers to cases of big creative processes or products that become known to all humankind and is closely related to the sociocultural definition an historical creativity. In contrast, little-c comprises of everyday problem-solving processes (Kaufman and Beghetto 2009). Recently mini-c and Pro-C creativity have been added to the group. Mini-c creativity describes personal learning process (Beghetto and Kaufman 2007) and pro-C creativity designates professional level creativity used in one's profession or at work (Kaufman and Beghetto 2009).

Overall creativity researchers see children as not being particularly creative. Children's creations rarely add any value to the knowledge adults produce and children also lack the knowledge to make and create artefacts that bring any value to humankind (Sawyer et al. 2003). However, at an early age children can learn to express creative behaviours in specific contexts (Hoicka et al. 2018, 2016). Usually these behaviours are original to the individual or a group of children but not necessarily important for culturally valued knowledge instead, enriching the general cultural environment. Therefore, children's creations would be categorised as little-c creativity or creativity on an individual level.

Creativity can be measured with different batteries of tests. Commonly researchers measure divergent thinking (Guilford et al. 1978; Wallach and Kogan 1965), which is the ability to generate as many novel items or ideas from one prompt such as a question about how a cup could be used. Usually methods to study divergent thinking rely heavily on the verbal abilities of the participants. However, the demand to study creativity in toddlers led to the development of novel non-verbal methods. One such example is the use of an "unusual box task" where a box full of holes and features is used to measure divergent thinking by examining the number of different ways children explore it (Bijvoet-van den Berg and Hoicka 2014). Studies using unusual boxes showed toddlers aged between 1 and 3 years old expressed divergent thinking (Bijvoet-van den Berg and Hoicka 2014; Hoicka et al. 2016, 2018). The divergent thinking scores between parents and their toddlers were even highly correlated (Hoicka et al. 2016). One of the reasons for the correlation might be attributed to cultural learning (see section 1.2) to which toddlers are constantly exposed when interacting with their parents or guardians.

Creativity measures mainly rely on measuring novelty, but for most of the innovations which are culturally transmitted, the usefulness or efficiency of the novel process or product is important (Tennie et al. 2014; Carr, Kendal, and Flynn 2015). Behavioural innovation may be defined as "a new, useful and potentially transmitted learned behaviour, arising from asocial learning (innovation by independent invention) or a combination of asocial and social learning (innovation by modification), that is produced so as to successfully solve a novel problem or an existing problem in a novel manner" (Carr, Kendal, and Flynn 2016; p. 11). It can be distinguished based on a novel end product or a novel process of behavioural innovation (Reader and Laland 2003). Innovation *sensu* product encompasses a new or a modified learned behaviour, which arises in a population. Innovation *sensu* process, is a process leading to a new or modified learned behavioural variant being introduced to the population. Innovation may be seen as a process generating a novel learned behaviour in an individual which does not solely originate from social learning and is not induced by the environment (Ramsey, Bastian, and van Schaik 2007). Innovation (whether by invention or modification) can be divided into three levels, where low-level innovation is unlearned "chance" innovation not repeated by individuals, mid-level innovation is useful for the individual and repeated (learned) by them and high-level innovation is useful at the population level as indicated by it being acquired (learned) by others (Carr, Kendal, and Flynn 2016). Innovation has long been regarded as important for understanding problem solving capabilities of individuals, but only in recent years has come to the attention of researchers in the field of social learning, in particular in children.

For the purpose of this thesis I define creativity as something novel (as measured by the comparison between different objects) (Table 1). Since I studied creativity in an open-ended task without the goal, children could not invent useful or appropriate solutions which would support the classical definition of innovation and in some cases creativity (Sternberg 2006; Sawyer 2012; Carruthers 2002; Carr, Kendal, and Flynn 2016). However, it can be argued creativity and its components such as originality and divergent thinking support innovation (Carr, Kendal, and Flynn 2016; Bateson and Martin 2013).

1.1.1 Children's innovation and creativity

Children are not seen as being particularly creative (Sawyer et al. 2003) or good problem solvers (Beck et al. 2011; Cutting, Apperly, and Beck 2011; Nielsen 2013; Tennie, Call, and Tomasello 2009) when creativity and innovation are seen as the production of useful behaviours or products. Many studies of children investigate their imagination and pretend play (Hoffmann and Russ 2012; Lillard 2017; for a review see Lillard et al. 2013) but not their capacity for originality within the domain of creativity that leads to novel products or processes. Instead of innovating, children predominantly learn from others (for a review see Wood, Kendal, and Flynn 2013). Furthermore, children aged 4 to 6 years do not succeed with problem solving tasks, such as the loop production task (Tennie, Call, and Tomasello 2009), the hook task (Beck et al. 2011) and the floating object task (Nielsen 2013). In the hook task a child is asked to retrieve a bucket from a tube and needs to manufacture a hook from a straight pipe cleaner to do so. The floating object task requires the child to fill a narrow tube containing a toy with water to retrieve it. The success rate of solving the former task was not predicted by divergent thinking or executive functions (Beck et al. 2016) suggesting divergent thinking and tool innovation are not necessarily linked. However, success was affected by perceived ownership of the tool and environment (Sheridan et al. 2016), prior experience with the tool (Whalley, Cutting, and Beck 2017) and clear affordances of the tool (Neldner, Mushin, and Nielsen 2017). Cultural background appears to have no effect on creativity/innovation as the floating object task was employed also in children from a non-Western background, where the rates of successful problem solving were the same as in Western children (Nielsen et al. 2014).

The hook task represents an ill-structured problem and it has been argued (Cutting, Apperly, and Beck 2011) that such tasks might not be suited to study children's innovation skills. Ill-structured tasks have a clear beginning state and a clear goal but the knowledge of the procedure to build and employ the tools is not known. If the tasks are indeed ill-structured they only show that children are unable to find the solution to a problem they are not faced with in day to day life. Children's lack of success at innovation tasks with a clear goal and limited solutions does not necessarily mean they do not have the ability to be creative and to innovate. In addition, study with adults showed that a well-defined problem can even dampen the performance on a creative task (Moreau and Engeset 2015).

Most of the studies measured innovation in isolation of the children's natural settings, such as being in a group and in informal learning environments, which are known to promote their problem solving abilities (Sheridan et al. 2016). It seems that when placed in the right context children can invent tools and express creativity (Hoicka et al. 2016; Reindl et al. 2016; Sheridan et al. 2016). The standard, previously mentioned, tasks to measure children's creativity were developed for the purposes of studying avian cognition (Weir, Chappell, and Kacelnik 2002; Bird and Emery 2009) and non-human primates (Mendes, Hanus, and Call 2007). They may be ill-suited to explore children's capabilities to innovate. However, showed children as young as 2 years old innovate and use analogous tools to the ones great apes use in the wild (Reindl et al. 2016). It is necessary to develop novel ways of measuring creativity and innovation in children.

Numerous studies have tried to test whether play fosters creativity in children (Garaigordobil 2006; Bateson and Martin 2013; Zosh et al. 2017). For example, for an entire school year, researchers exposed children aged 10 to 11 year old to different activities that supported verbal, dramatic, graphic-figurative and plastic-constructive creativity. Children played with the games that supposedly fostered creativity in different domains in small groups. At the beginning and at the end of the exposure to the playful games, children that were exposed to the play treatment had to complete a battery of tests that measured their ability to produce novel ideas (fluency), their ability to change ideas and thoughts (flexibility) and their ability to come up with numerous unique solutions (originality). Children who were regularly exposed to the playful games in groups performed better than the control group that was not exposed to any treatment. There were also no differences between the sexes of children in relation to their scores on creativity tests (Garaigordobil 2006).

Children are heavily reliant on social information (Legare and Nielsen 2015; Dean et al. 2012; McGuigan et al. 2017) and may use copying as a strategy independent of the value of the social information (Carr, Kendal, and Flynn 2015; Turner, Giraldeau, and Flynn 2017; Flynn, Turner, and Giraldeau 2016; Clay, Over, and Tennie 2018). For example, the majority of 4 to 5 year old children (90%) imitated, rather than innovated (by modification) the solutions to solve a puzzle box and extract a reward, often despite a very low efficiency of the demonstrated solution (Carr, Kendal, and Flynn 2015). However, when the demonstrated method was inefficient the innovation attempts in children were greater relative to imitation, in children aged 6 to 7 years compared to 4 to 5 year olds (Carr, Kendal, and Flynn 2015). Children aged 3 to 5 years more likely ask for a demonstration when faced with a novel task (Turner, Giraldeau, and Flynn 2017). Furthermore, studies indicate that innovation by invention is more demanding for younger (4 years old) children than for older (8 year old) children (Nielsen 2013; Beck et al. 2011).

In daily life children interact with, learn from, and copy the behaviour of their peers as well as adults (Haun, Rekers, and Tomasello 2014; Zmyj et al. 2012; McGuigan and Robertson 2015). We can see that younger children copy more often than innovate (Carr, Kendal, and Flynn 2016), whereas older children start to use and develop different skills that facilitate behavioural flexibility and innovation (Gopnik et al. 2017; German and Defeyter 2000; Carr, Kendal, and Flynn 2015). The vast majority of novel creations and useful innovations at the population level are delivered and modified through imitation and tradition (Sawyer, 2012, p. 28). It can be argued that social learning together with innovation form the pillars for cumulative cultural evolution which leads to the development of the diverse cultural artefacts seen in humans (Legare and Nielsen 2015; Derex and Boyd 2015).

1.2 Social and cultural learning in children

Children are very good at social cognition tasks (Herrmann et al. 2007; Boyd, Richerson, and Henrich 2011), they excel in learning from others (Dean et al. 2012). A predominant type of learning is social learning. Social learning, defined as, "learning that is influenced by observation of, or interaction with, another individual, or its products" (Heyes 1994, p. 207), is widespread in children (Harris 2012). Social learning in developmental and comparative psychology is also encompassed under the term cultural learning (Tomasello, Kruger, and Ratner 1993).

Cultural learning can be divided into three categories of learning: imitative learning, instructed learning and collaborative learning (Tomasello, Kruger, and Ratner 1993). Overall through cultural learning children learn how to use objects around them (Hopper et al. 2010; Demps et al. 2012; DiYanni, Nini, and Rheel 2011; Flynn and Whiten 2008) and behavioural norms in specific contexts (Over and Carpenter 2013; Rakoczy et al. 2009; Wang, Williamson, and Meltzoff 2015). However, cultural learning through observing does not typically happen without communication being present and conversations among children and adults support informal learning in children (Csibra and Gergely 2011; Paradise and Rogoff 2009). Children follow certain social learning biases (Wood, Kendal, and Flynn 2013b) and employ specific strategies to solve problems they are faced with. For example they follow the majority (Burdett et al. 2016; Corriveau and Harris 2010; Evans et al. 2018) and conform with others (Morgan, Laland, and Harris 2015; Haun and Tomasello 2011; Haun, Rekers, and Tomasello 2014; Schillaci and Kelemen 2014). Social learning is an important component of a child's development, through which they learn how things are, or should, be done (Göckeritz, Schmidt, and Tomasello 2014), and whom to trust (Harris 2012). Arguably social learning limits innovation (Carr, Kendal, and Flynn 2015; Whiten and Flynn 2010; Reindl and Tennie 2018; McGuigan et al. 2017; Derex and Boyd 2015) and direct instruction (a form of active social learning) constrains children's exploration of objects (Bonawitz et al. 2011; Van Schijndel and Raijmakers 2016). The relationship between creativity defined more broadly, as something that is novel, and social learning is not clear.

Social learning and teaching enables the transfer of knowledge, skills, practices, and other forms of non-genetic information from individual to individual. It comprises different processes. I will focus on imitation and emulation because these are the most common methods of acquiring behaviour from demonstrators or peers. Imitation is attributed when a participant copies the sequence of actions performed by a model to fulfil a desired goal (Whiten et al. 2009; Want and Harris 2002; Hoppitt and Laland 2013). When an observer copies the final goal of a model's actions, but not the sequence of actions that they perform, emulation is considered to have occurred (Hoppitt & Laland 2013). More specifically, an observer acknowledges that a specific result may be achieved, but may accomplish it by different means (Want and Harris 2002). Research on chimpanzees indicates a greater reliance on emulation as opposed to imitation, whereas children usually imitate (Horner and Whiten 2005; Want and Harris 2002).

The imitation abilities of children are already present in new-born infants who have been found to copy adult facial gestures (Meltzoff and Moore 1983). Infants are also attracted to, and can memorize, other peer and adult gestures and actions. For example, 14-month old infants imitate demonstrations and actions presented to them by adults 24 hours prior (Meltzoff 1985, 1988). One of the tasks used included an adult demonstrator who turned a light-box on by touching it with their forehead (Meltzoff 1988). Toddlers in the study copied the action even though they could emulate the action and turn on the light-box by touching it with their hand. This shows children heavily rely on imitation and even 'over-imitate' (Whiten et al. 2009; Clay, Over, and Tennie 2018; Moraru, Gomez, and McGuigan 2016; Lyons, Young, and Keil 2007; Keupp, Behne, and Rakoczy 2013; Evans et al. 2018; Lyons et al. 2011). Alternatively, it has been shown that when children clearly understand the reason why adult demonstrators do not use their hands to turn on a light-box (their hands are occupied) since the majority of children emulate, by using their hands to turn on the light box (Gergely, Bekkering, and Király 2002). Furthermore, when infants of 12- and 18-months are shown an action where different goals are presented they adjust their copying strategy depending on the goal of the tasks they are presented with (Carpenter, Call, and Tomasello 2005). For example, when demonstrated an action with a clear goal of putting a mouse in a toy house, children did not imitate the way the mouse was moved by the researcher. However, when the mouse did not have a clear end destination, a toy house, and only expressed random movements, children understood these moves as an important behaviour and copied them (Carpenter, Call, and Tomasello 2005). Throughout the second year of development children employ different strategies of copying adult demonstrators (Nielsen 2006).

Children's social learning is susceptible to many biases (for a review see Wood, Kendal, and Flynn 2013). For example, in a formal experimental environment (school) children are more likely to imitate an older than more knowledgeable model (Wood, Kendal, and Flynn 2012, 2013a). Even as toddlers (15-month of age), children are more likely to copy adult videotaped actions than the actions of a 24-month-old toddlers (Seehagen and Herbert 2011). When three- and four-year-olds receive reliable information on how to label novel objects from a child and adult, they preferentially copied the adult model (Jaswal and Neely 2006). However, children also preferentially copy others depending on the context. Infants (14-months of age) copy other same aged infants when they imitate gestures (Zmyj et al. 2012), and three-year-olds copy their peers when it comes to play, such as putting a teddy bear to bed (Ryalls, Gul, and Ryalls 2000). Whereas if adults perform playful behaviour and express irrelevant behaviour, children (4 – 5 years old) prefer to copy other children (Nielsen, Cucchiaro, and Mohamedally 2012).

Diffusion design studies (Mesoudi and Whiten 2008), where a model demonstrates a behaviour within a naïve group, in children are not common. A review (Flynn and Whiten 2010) recognised five diffusion studies with young children. In all studies, puzzle boxes that children had to solve were used. Research mainly focused on factors that influenced the fidelity of transmission across generations of children. In children aged between 2 to 5 years, factors such as age, social standing and personality traits all influenced social learning, but the child's sex did not (Flynn and Whiten 2012). The type of knowledge transmitted also impacts on the fidelity of its transmission. For example social information in the form of gossip is transmitted more often than formal knowledge in 10 to 11 year old children in comparison to general knowledge information such as the number of people who speak Chinese (McGuigan and Cubillo 2013).

Studies combining social learning with innovation by modification have mostly explored the ratchet effect in humans (Tomasello 1999) that leads to the cumulative improvement of cultural traditions and artefacts. It is understood that human cultural traditions grow through the modifications of previous knowledge operating as a ratchet. Ratchet prevents a slippage of information and through social learning, leads to improvements over time. When a ratchet effect is present the performance on a task improves over the course of the transmission of a behaviour between individuals in a chain (Caldwell and Millen 2008a; Caldwell and Millen 2008b). However, in these studies using transmission chains, typically only one participant builds whilst others in the group observe before they attempt the task with new observers present. This carries on along the chain until the last participant in the chain finishes the task (Horner et al. 2006). For example adults were asked to make paper airplanes that flew as far as possible (Caldwell and Millen 2008a, 2008b) and build the tallest spaghetti and clay towers (Caldwell and Millen 2008a, 2008b, 2010; Reindl and Tennie 2018). In adults, the ratchet effect was clearly present since the length of flights increased across generations and the spaghetti towers became higher (Caldwell and Millen 2008a). It has also been shown that the artefacts in the same chains are more similar than in others appearing as separate cultures.

Learning in humans is facilitated not only by inadvertent social learning, but also by teaching (active social learning), language, and prosociality (Csibra and Gergely 2009; Tomasello, Kruger, and Ratner 1993). Cultural learning therefore also includes instructional learning and learning through cooperation (Tomasello, Kruger, and Ratner 1993; Legare and Nielsen 2015). Studies have shown that prosocial behaviours in humans may underpin cultural transmission processes (Moll and Tomasello 2007; Herrmann et al. 2007; Dean et al. 2012). The aforementioned theory and findings cannot be excluded when researching children in naturalistic environments. In everyday life children express diverse behaviours and learn in many different ways (Gopnik 2012; Legare 2017), besides they are sometimes also actively thought (Kline 2014; Fogarty, Strimling, and Laland 2011; Ashley and Tomasello 2001).

Children are attentive towards signals from other humans from an early age this is incorporated in the term 'natural pedagogy' that describes communicative adaptation to transfer generic knowledge between individuals (Csibra and Gergely 2009). Children in all cultures learn through similar processes and not necessarily just through direct teaching often used in Western educational systems (Demps et al. 2012). Apprenticeship, direct demonstration, and delivered feedback may be interpreted as part of the teaching processes, where humans communicate when they demonstrate an action (Lancy 2017).

1.3 Ecological validity

The majority of studies of children's innovation, presented in this brief review, use tasks in which children modify and, in some cases, innovate tools (Beck et al. 2011; Tennie, Call, and Tomasello 2009; Nielsen 2013; Reindl and Tennie 2018) or create better, more efficient solutions to extract rewards (Carr, Kendal, and Flynn 2015). All these studies heavily rely on children's problem-solving abilities and not on the generation of creative novel solutions. However, all these studies still used puzzle boxes, where children were faced with a problem they needed to solve. For example, usually children are asked to retrieve something from a box using a set of tools (Carr, Kendal, and Flynn 2015; Flynn 2008; Whiten and Flynn 2010) or discover solutions through interacting with the box (Dean et al. 2012; McGuigan et al. 2017). Researchers expose children to specific tools they can use and from which they can find and create solutions but do not leave them to freely explore and solve an open-ended problem which might have numerous solutions. However, divergent thinking in children has been previously studied in a more open-ended way with tasks in which children need to suggest or perform as many uses as they can for a specific object (Bijvoet-van den Berg and Hoicka 2014; Hoicka and Bijvoet-van den Berg 2013).

I was interested in exploring creativity based on its novelty rather than problem solving. The reason for children's low ability to innovate might be that they were asked to solve a goal oriented task, which restricts children to just correct one (Beck et al. 2011; Nielsen 2013) or in some cases a few solutions (Carr, Kendal, and Flynn 2015). Instead of a goal-directed task I propose to explore the connection between creativity and social learning is with an open-ended task and exposing children to different contexts in an informal learning environment. Furthermore, my aim was to find a task to study children's creativity which was similar to what children would encounter in their everyday life. For example, using mediums such as toys and blocks. Specifically, wooden blocks are used to teach children physics and to foster the development of their spatial cognition (Casey and Bobb 2003; Casey et al. 2008; Ferrara et al. 2011; Newman, Hansen, and Gutierrez 2016) and are often present in the home as toys/games as well as at nursery or school. Some previous research even showed that how children play with blocks is a good predictor for children's later mathematical skills (Wolfgang, Stannard, and Jones 2001; Stannard et al. 2001). Wooden blocks also offer infinite possibilities for building and this open-endedness offers a good view into children's creativity in a naturalistic setting with an ecologically valid task.

Most of the research investigating social learning has been carried out in controlled environments, such as psychological laboratories and formal educational context. There is a considerable gap within the cultural evolution and psychological literature of research done in naturalistic group contexts. Social learning and skills required to innovate are mainly studied with dyadic experiments, where one participant observes a demonstrator and tries to solve a puzzle box on their own (Wood, Kendal, and Flynn 2012, 2013a; Nielsen, Moore, and Mohamedally 2012; Buchsbaum et al. 2011; Nielsen 2013). Children are usually constrained and can attempt the task just a few times, which does not represent the naturalistic contexts that children find themselves in (Reindl and

Tennie 2018). Studies that have used naturalistic settings often involve children in open diffusion (McGuigan et al. 2017; Whiten and Flynn 2010; Flynn and Whiten 2012; Dean et al. 2012) and not in limited dyadic experiments. Open diffusions closely resemble a naturalistic setting because children can freely interact with one another which encourages a natural learning context. However, the task is usually set up so only one individual can interact with it at once.

Behavioural observations in naturalistic contexts of children imitating peers and children imitating adults showed children copy peers in playful behaviours and adults when the context is teaching, directing or discipline (Grusec and Abramovitch 1982). In another naturalistic experiment children, aged between 4 to 12 years, over-imitated or copied irrelevant causal actions when freely interact with a puzzle box after seeing a demonstrator solve it (Whiten et al. 2016). With open diffusion experiments the transfer of knowledge (Flynn and Whiten 2012, 2010; Flynn 2008), cumulative problem solving (Dean et al. 2012) and the existence of a ratchet effect (McGuigan et al. 2017) have been studied in groups of children in nurseries. Researchers could explore how visitors interact and tackle specific problems, that can be solved through innovation or social learning, in a similar way to a study performed in the past in a Zoo (Macdonald and Whiten 2011; Whiten et al. 2016).

Play is an interesting aspect of children's behaviour which can be seen as a creative process (Hewlett and Boyette 2012). Play has been explored in both controlled and natural contexts such as in nurseries. For example, children copy peers in the context of play (Ryalls, Gul, and Ryalls 2000; Nielsen, Cucchiaro, and Mohamedally 2012; Zmyj et al. 2012) whereas they are more likely to rely on other social learning biases (eg. copying adults) in less playful contexts such as solving a task by innovating a novel tool

(Beck et al. 2011). In the latter example the children view the adults as proficient users whereas in the first example in the context of play or interacting with a toy, a peer is regarded as the proficient model (Carr, Kendal, and Flynn 2015). Exploration of play show how important environmental and social context can be for learning.

Studies of prosocial behaviours in diverse contexts and societies show learning takes many different forms and the transmission of knowledge across groups has different paths (Kendal et al. 2018; Wood, Kendal, and Flynn 2013b; Terrashima and Hewlett 2016). With typical dyadic experiments and transmission chain experiments we are limited to studying only one way of transmission. However, research in contemporary hunter-gatherers has shown that knowledge can be transmitted in different directions and is susceptible to diverse social learning strategies (Hewlett 2016). For example, when observing play in children from Aka forager community, imitation of traditional activities does not happen when adults are present but more commonly when they are absent (Boyette 2016). During this time in groups of children cultural transmission happens through play. While children from WEIRD populations are mostly studied in the limited environments of laboratories, nurseries and schools, science centres offer a venue to explore children's behaviours in an informal learning context where behaviours might be more similar to those recorded in small scale societies, such as children playing independently without the presence adults.

There are undoubtedly benefits of performing experiments in the laboratory which offer high control and with it high, experimental validity. Experiments enable researchers to "re-run history multiple times, manipulate variables, and accurately record behaviour" (Mesoudi 2011, p. 159). During the process of development children inhabit specific environments constructed for them by adults, although they may construct their own learning within this (Kendal 2011; Flynn et al. 2013). Human culture does not develop in a vacuum and is highly influenced by factors such as environment and social factors (Miton and Charbonneau 2018). Small-scale experiments are not sufficient to study long term changes in cultural evolution. Therefore we need to follow and perform observations in environments that enable research in the real world over a larger timescale (Mesoudi 2011). Children's social learning and creativity need to be studied in more natural settings which offer higher ecological validity than do experimental settings. The experimental design usually limits children and guides them towards a limited or specific solution to tasks, whereas in naturalistic settings children are not limited but allowed to freely explore and express all behaviours within a certain domain. One of the social contexts where social learning and creativity can be explored are informal learning environments. Science centres offer great opportunities to study children because they provide a place where they can be studied within a specific informal learning contexts (Callanan 2012; Jipson and Sobel 2015).

1.4 Behavioural studies in science centres

In science centres innovative behaviour is encouraged (Gartenhaus 1997), and various generations come together, explore the exhibits, and often learn something new (Falk 2009; Jipson and Sobel 2015). Children gain knowledge and learn in everyday social contexts. These contexts include interactions within the family and within diverse communities in a form of informal learning whereby they learn from others through observing and participating rather than being actively taught (Paradise and Rogoff 2009). One form of informal learning children experience in some societies, is that available in science centres, which present natural social contexts that children are exposed to such as playing and interacting with peers. Visitors in science centres spontaneously express various social learning processes such as imitation. Such centres are thus perfect places to study how knowledge and skills are generated and culturally transmitted via social learning (Callanan 2012).

Previous studies in science centres mainly focused on how well child and adult visitors learn about science with a particular focus on the STEM subjects (Jipson and Sobel 2015). For example, researchers are interested in which type of exhibits foster explorative behaviours in children (Van Schijndel, Franse, and Raijmakers 2010; Sandifer 2003; Callanan, Martin, and Luce 2015). A study in a science centre explored how being exposed to an analogical model of a tower before building with specific blocks affected how children solved the problem of building a stable tower. Children aged from 6 to 8 years old, that were exposed to a stable model of a tower, emulated the product they had seen and used the methods in their own build (Gentner et al. 2016). This study indicates that being exposed to a solution enabled children to emulate the solution. However, children were not only exposed to passive teaching and observation since parents built with them and may have taught them about the spatial demands of the task. This study incorporated different types of learning through exposing children to conversation, teaching and exposing them to the model of a tower but did not investigate their creativity.

To sum up, science centres offer a good opportunity to study different types of cultural learning in children within a context that is more natural to children, besides informing people about social learning studies and studies of innovation in children.

Table 1: Table of theoretical and operational definitions of the traits I explored within the research presented in this thesis

Term	**Theoretical definition**	**Operational definition**
social learning	"learning that is influenced by observation of, or interaction with, another individual, or its products" (Heyes 1994, p. 207)	either direct instructions, or scaffolding type instructions (Chapter 3), building at the same time as another participant or seeing the final structure of another participant (Chapter 4)
asocial (individual) learning	learning through individual experience and trial and error learning	not receiving any instructions (Chapter 3) or not being able to see anyone else building at the same time or their final structures (Chapter 4)
cultural learning	a combination of social (imitative) learning, instructed learning and collaborative learning. (Tomasello, Kruger, and Ratner, 1993)	either direct instructions, scaffolding type instructions (Chapter 3), building at the same time as another participant, seeing the final structure of another participant (Chapter 4)
cooperation	individuals helping each other	working together on the same parts of a structure (Chapter 5)
collaboration	working together to reach the same goal	working together but on different parts of a structure (Chapter 5)
innovation	"a behavioural innovation is new, useful, and potentially transmitted learned behaviour, arising from asocial learning (innovation by independent invention) or a combination of asocial and social learning (innovation by modification), that is produced so as to successfully solve a novel problem or an existing problem in a novel manner" (Carr, Kendal, and Flynn 2016, p. 11)	innovation was not directly measured with the tasks used in this thesis; I have measured creativity which leads to innovation (Carr, Kendal and Flynn, 2016) through originality (see operational definition of creativity)
creativity	"ideas, behaviours, or products that are both novel and valuable" (Picciuto and Carruthers 2014, p. 199)	originality as measured by the comparison of two structures by many raters (Chapters 3, 4, 5)
ecologically valid context	"observations in environments that enable research in real world" (Mesoudi 2011)	two aspects ecologically valid task (building blocks task) and ecologically valid environment (science centre)

1.5 Main aims of the thesis

In this introduction I presented studies of cultural learning and innovation in children. These studies, mostly performed in controlled contexts (such as laboratories and school class rooms), have shown that children at the age of 4 years learn from others and can think divergently. For example, toddlers as young as 1 year old think divergently (Hoicka et al. 2016, 2018) and when up to 2 to 3 years old, can innovate analogous tools to those that chimpanzees develop in nature (Reindl et al. 2016). By the age of 4 children create novel objects and, in some cases, innovate tools in controlled experimental settings. Children develop basic spatial cognition skills with blocks by the age of 4 (Reifel 1984) therefore any tasks with building blocks should not be too cognitively demanding for them. In contrast, only a small number of studies have explored social learning in detail in less controlled and thus potentially more "natural" contexts (Flynn and Whiten 2010; Whiten et al. 2016; McGuigan et al. 2017; Dean et al. 2012). These studies provide ecologically valid contexts since they use open diffusion experiments in groups of children in nurseries, schools and zoological gardens. The findings corroborate with those of laboratory studies as well as provide new questions which can be studied in detail in more controlled settings. For example, counter to expectations, a study of over-imitation in children and adults in the presence of a demonstrator recorded over-imitation in all participants irrespective of their age in the real-word context in a Zoo (Whiten et al. 2016). When exploring the processes which underlie human cumulative culture and abilities to solve more complex tasks with additional stages of difficulty, children have been found to not exclusively rely on social learning but use a diverse set of prosocial behaviours and instructions to retrieve the reward and in some cases, even share the reward (Dean et al. 2012).

In this thesis I present studies with a wide ranges of ages. Children were aged between 4 to 12 years old. First, I used this age group since by this age children are proficient block builders and have developed spatial skills they need for the tasks used in my research (Reifel 1984). Second, by the age of five children might preferably solve problem solving tasks individually (Turner, Giraldeau, Flynn 2017), however, they do over-imitate in real word contexts (Whiten et al. 2016). Third, social learning strategies are usually studied in toddlers and children in early classes of primary school (for a review see Wood, Kendal, Flynn 2013), while the literature on children in the wider age range between 4 to 12 is scarce. Three to fourteen year old children from diverse societies increase the probability of expressing non-costly prosocial behaviours as they mature (House et al. 2013). Fourth, research has shown young children are creative (Hoicka et al. 2016, 2018) but they do not necessarily excel in tasks that test for their innovation skills. Only children at age 7 to 8 are on the same level as adults of accomplishing the hook task (Beck, Apperly, Chappell, Guthrie, & Cutting, 2011; but see Sheridan, Konopasky, Kirkwood, & Defeyter, 2016) and 4 year old children are not successful at solving the floating object task (Nielsen 2013), but 3 year old children are proficient at innovating analogous tools to those made by chimpanzees in nature (Reindl et al. 2016). Fifth, I chose a wide age range because since the zones in the science centre were aimed at all primary school aged children. Hence, science centres offer a great venue to conduct research with a wide age.

The main goal of the three studies I performed in this thesis was to explore cultural learning in children through the use of an open-ended task in a relatively informal learning setting (a science centre). I was particularly interested in the interplay between cultural learning and creativity in an open and natural environment children would find themselves in. In science centres exhibits are spread across open floor design spaces and children and adults can interact with them freely without any restrictions. Arguably for western children science centres present a common and natural environment and are thus a perfect place to conduct behavioural and developmental research with children (Callanan 2012). The thesis had the additional applied aim of informing science centre practitioners about the impact of instructions on children's behaviour in the science centre and how to design exhibits to promote cooperation, cultural learning or asocial (individual) learning.

1.6 Thesis outline

In this thesis I present research that has been performed in an informal learning environment, a science centre, with children between 4 to 12 years old. In Chapter 2, I present the science centre where I performed the studies presented in this thesis. This is followed by a description of the novel open-ended tasks using two types of wooden building blocks and a description of a novel exhibit which supports research of social learning in the science centre and collects data automatically without the need of a researcher being present. Furthermore, I introduce a novel web application to collect similarity ratings from pairs of images. Last, I present Bayesian statistical analysis which has been used throughout the thesis. After the general methods (Chapter 2), follow empirical chapters. Specifically, in Chapter 3 I investigated how direct instructions, scaffolding and no instructions impact children's exploration, building block behaviours and the originality of their final structures. In Chapter 4 I tested the effect of social and asocial learning on the originality of structures children built. In

Chapter 5 I performed an explorative study of children's behaviours in pairs and groups. I tested whether any of the behaviours they expressed in groups, such as instructing, discussing, cooperating and collaborating effected the originality of structures pairs and groups of children built. In the general discussion (Chapter 6), I discuss the findings of the studies in relation to the social learning and innovation literature in children. I touch upon performing research in naturalistic contexts and the use of Bayesian statistical analysis. At the very end I present benefits and limitations of these studies and present some ideas for the future.

2 General Methods

2.1 Research Context

The research presented in this thesis was conducted in a science centre, Life Science Centre, Newcastle, UK. The Life Science Centre (LSC) is a public science exhibition, with theatre spaces and teaching laboratories for schools, in Newcastle upon Tyne. It is located within the Centre for Life, a 'science village' comprising medical research laboratories, National Health Service clinics, commercial businesses and the public science centre. It is operated by the International Centre for Life Trust and the rental income from the university and business tenants provides the majority of the operating funding for the science centre. This, combined with admission fees, makes the LSC unique as an operationally self-sufficient science centre.

Figure 1: The Life Science Centre

The majority of visitors in the LSC were from the North East of the UK (Table 2). At the time the study took place, the LSC had five zones: Curiosity Zone, Brain Zone, Experimental Zone, Young Explorers Zone and a zone with a temporary exhibition. Besides the educational zones the centre has a planetarium, a 4D motion ride and a science theatre. The admission fee for adults was £15 and for children (5 to 17 years old) £8, children 4 years and younger had free admission. I conducted the research in the Curiosity Zone (2.1.2) and in the Brain Zone (2.1.3).

Table 2: LSC visitor's origin, sex and profile

Visitor origin	[%]	Visitors sex	[%]	Visitor profile	[%]
Newcastle	35	Female	68	Family groups (with children of ages up to 12 year old)	70
Rest of Tyne and Wear, Durham and Northumberland	55				
Cumbria, Borders, North Yorkshire	6				
Rest of UK	3	Male	32	Schools	14
Overseas	1			Other audiences (adults and teenagers)	16

2.1.1 Some of the previous studies done at the Life Science Centre

The LSC has prior experience collaborating with academics. Researchers from York University, Durham University and Newcastle University have performed studies on children's behaviours in groups (Over et al. 2018) and their tool innovation (Evans et al. 2018; Sheridan et al. 2016).

2.1.2 Curiosity Zone

The Curiosity Zone was opened in May 2012. Its main aim is to foster curiosity, exploration and experimentation in the centre's visitors. The zone is comprised of open-ended interactive exhibits. Although there are no direct instructions on how to interact with exhibits, the LSC practitioners added guidance in the form of open questions in response to feedback from visitors (Andy Lloyd pers. comm.). The Zone is aimed at both children and adults.

2.1.2.1 Study One in the Curiosity Zone

In the first study (Chapter 3) I used an exhibit already present in the Curiosity Zone at the LSC to compare the effect of direct instructions, scaffolding, and no instructions had on children's creativity. The block exhibit consisted of numerous wooden blocks of different lengths (Figure 3). The length corresponded to the number of cubes constituting the block. These cubes were connected by rope that ran through the middle of the block. The rope and special holes in the cubes enabled the cubes to be rotated relative to each other such that blocks could form different shapes. The exhibit was designed to foster spatial awareness, motoric ability, creativity and exploration.

Figure 2: First study place behind partitions dividing the space for the wooden blocks exhibit on the science centre floor

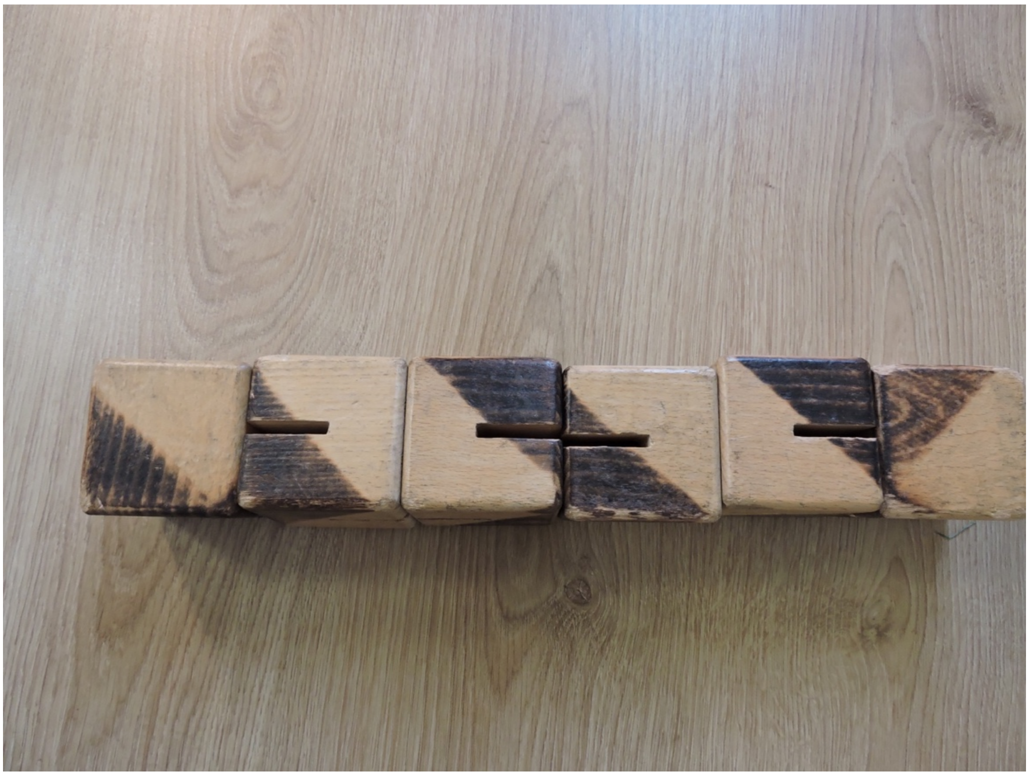


Figure 3: Blocks used in the first study (Chapter 3)

2.1.3 Brain Zone

The Brain Zone was opened in April 2016 after I had conducted the first study (Chapter 3) presented in this thesis. The Brain Zone constitutes of exhibits which introduce visitors to behavioural experiments with which scientists explore the workings of the human brain and mind.

2.1.3.1 The Interactive Research Pod

One of the exhibits within the Brain Zone was the Interactive Research Pod (IRP) which I was involved in developing as part of a team of academics from Durham University and museum practitioners from the Life Science Centre (Rudman et al. 2017; Bailey-Ross et al. 2018; Kendal et al. 2016). The exhibit presented a way of studying social learning in children and enabled researchers to collect data automatically. It supported diverse studies of social learning in children in a natural environment. I used the IRP exhibit in Study 2 to compare children's creativity when presented with a building block task in social and asocial learning conditions (Chapter 4). In Study 3 I used the IRP exhibit to research creativity in pairs and groups of collaborating children, again using the building block task (Chapter 5).

The Interactive Research Pod (Figure 4) offered a creative activity where children were asked to make a structure out of wooden blocks. The exhibit comprised of a table which was divided into three separate building stations by partitions. The partitions could be opaque, transparent or removed completely which enabled me to collect data under three different conditions: individual (asocial learning), social learning and cooperation. Under individual (asocial) conditions a single user built on their own and was unable to see other users since the partitions mounted at the IRP were opaque. Under social conditions single users built on their own at the station but were able to see others building at the other two stations through the transparent partitions. Under cooperation conditions the partitions were removed entirely, and users could build together on a table that was not divided into stations.

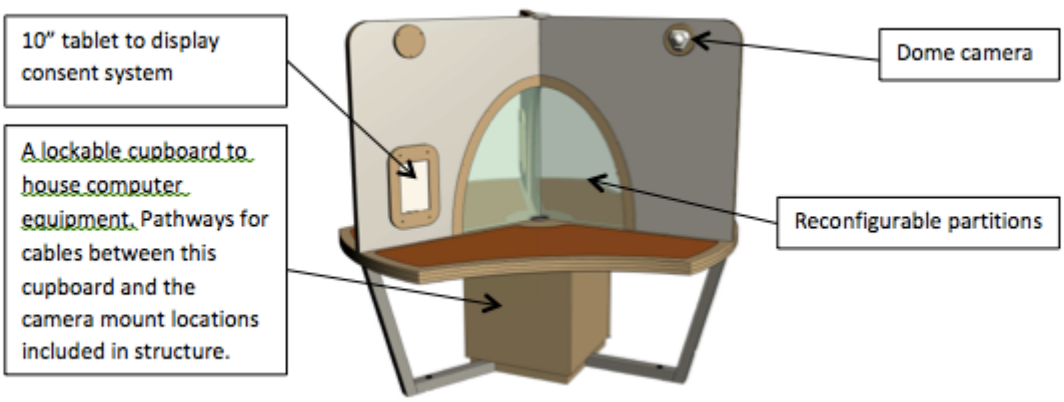


Figure 4: The diagram of the Interactive Research Pod

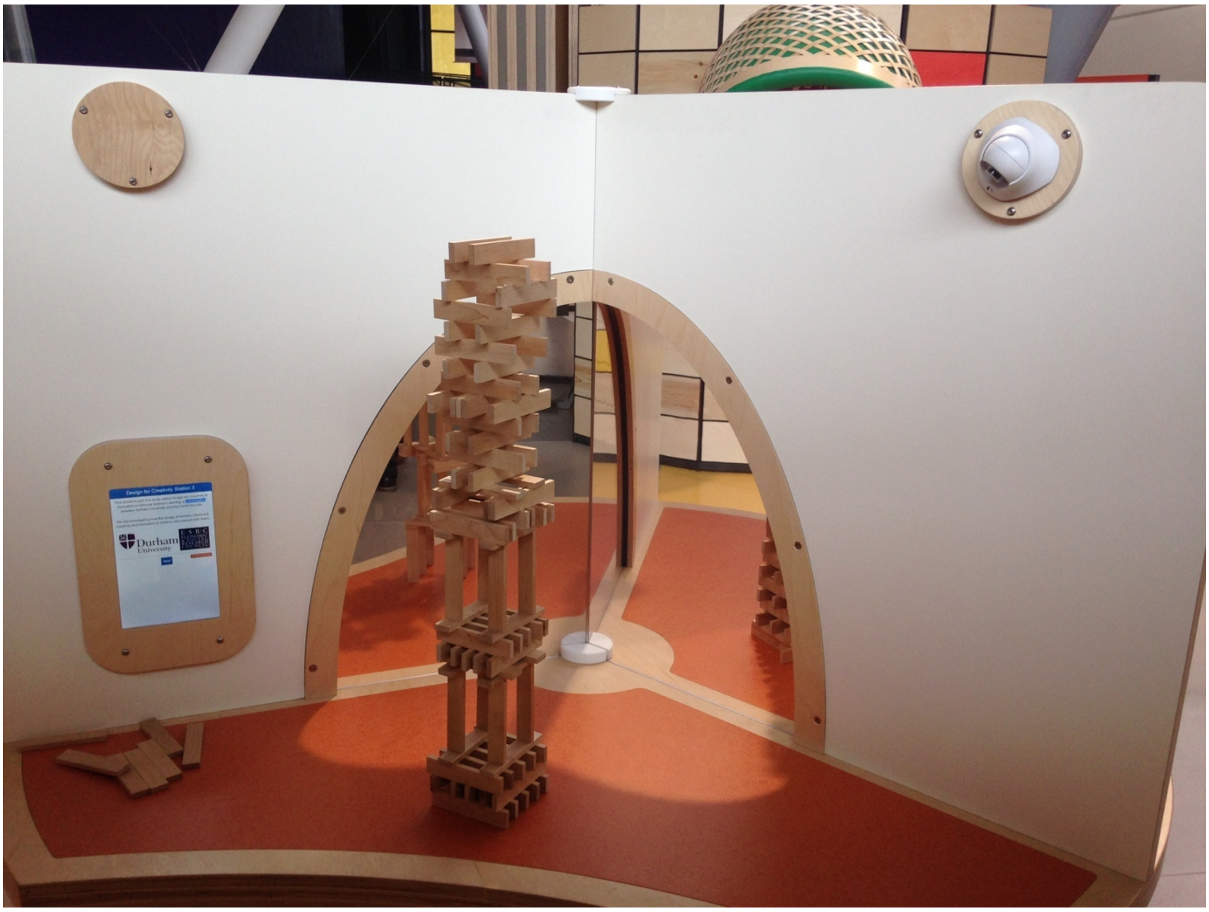


Figure 5: The Interactive Research Pod

Cameras and tablet computers were mounted onto the partitions in order to record the participant's behaviour, gather general information about them and gain their assent to participate in the studies. Consent forms were offered on an open source survey system, Limesurvey (http://www.lime-survey.org). The questions presented on the consent forms adhered to the guidelines for internet facilitated research given by the British Psychological Society and had been approved by the ethical consent board of Durham University. Additionally, there was a camera mounted over the entire IRP to record people interacting with the exhibit from above. The date and time that were present on the video recordings (Figure 6) were used to pair with the data collected on the consent forms which had time stamps. This ensured the data collected for analyses were from individuals that gave assent and the assent form data corresponded with the correct image of the structure built at the IRP. The IRP had Internet Protocol (IP) cameras which were connected to a Network Video Recorder (NVR) stored in a locked cupboard. The data collected by the consent forms and cameras were automatically stored on a computer with a 3 TB hard drive that was hidden under the table in a secured cupboard. The data were later transferred through a safe local network to servers within the museum where I could access them. All wires were hidden away within the wooden structure of the exhibit.

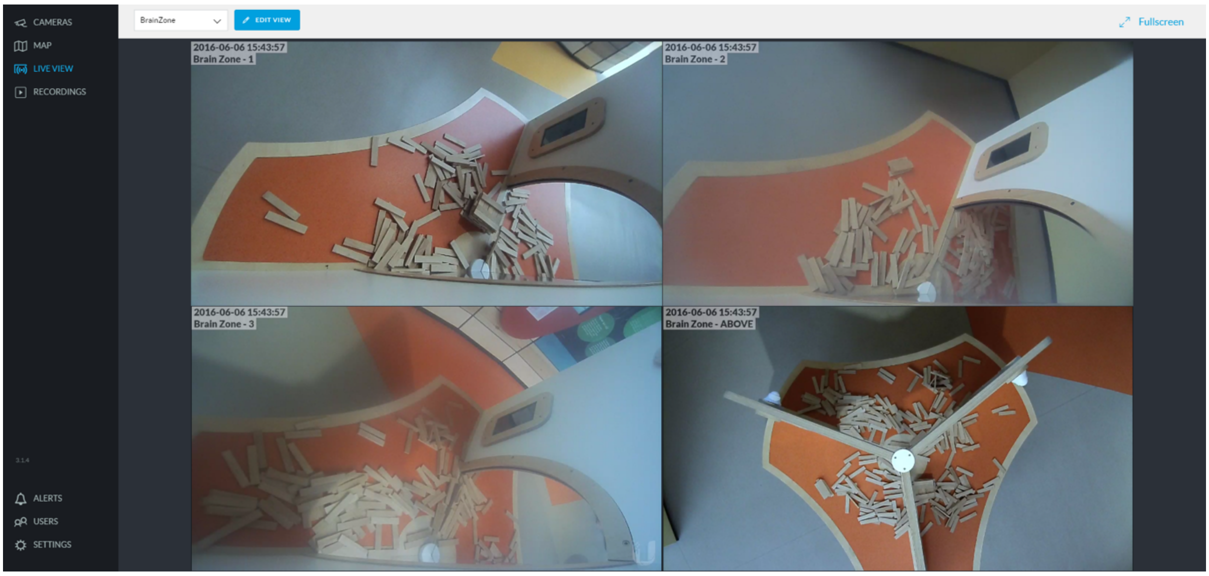


Figure 6: The screen shot from the videos the IRP collected automatically

At the Interactive Research Pod children were asked to build with up to 100 wooden blocks which were all the same shape. These blocks were Keva planks (www.kevaplanks.com, Figure 7) and were made out of maple, produced in the United States and frequently used in science centres, museums, formal education and libraries around the world. All blocks were the same light brown wood coloration and each block was 6.35 mm thick, 19.05 mm wide and 114.3 mm long. Children and adults could create any structure they wanted by simply stacking blocks together to create any structure they wanted. The wooden blocks thus afforded a huge range of building possibilities making them an ideal medium for open-ended tasks to explore children's creativity.

Figure 7: A single KEVA plank (source: https://en.wikipedia.org/wiki/File:KevaPlank.jpg)

2.2 Data collection

2.2.1 Recruitment of child participants in the science centre

In the first study (Chapter 3) I actively recruited participants by approaching and talking to their parents when they entered the exhibit area. The children who participated in the study received a sticker as reward for their participation.

In the second study (Chapter 4) the data collection was performed automatically, and participant recruitment was not required since the exhibit was open for anyone to interact with, and could record consent and behaviour automatically.

In the third study (Chapter 5) I was assisted by a research assistant (Guy Lavender-Forsyth) since I was investigating cooperation in children and by working together we could recruit more than one participant simultaneously. We actively recruited participants by approaching and talking to their parents when they entered the exhibit area. The children who participated in the study received a sticker as a reward for their participation.

2.2.2 Measuring behaviours in children

In studies 1 (Chapter 3) and 2 (Chapter 4) I coded all the behavioural data from videos. Whereas in study 3 (Chapter 5) a research assistant and I coded children's behaviours in groups as they occurred.

2.2.3 Recruitment of adult participants for rating of the structures children built

For the three studies presented in this thesis I recruited groups of students from the Anthropology Department at Durham University and adults through the online recruitment service Prolific (www.prolific.ac; Palan and Schitter 2018). I asked them to rate whether the structure on the image corresponded to what the child said they had built and to rate the similarity between pairs of structures children built.

2.2.4 Rating creativity using pairs of images of the structures children built – development of the application

Since creativity and innovation can be defined in different ways (Runco and Jaeger 2012; Ivcevic 2009; Stein 1953; Carruthers 2002; Runco 2014) and are traits which are hard to quantify objectively (Caldwell and Millen 2008b; Reindl and Tennie 2018), I developed a method to rate pairs of structures children built based on their similarity. The method is based on the similarity ratings used in studies of cumulative culture in adults (Caldwell and Millen 2008b) and has been recently been used in studies with children building spaghetti towers (Reindl and Tennie 2018). In both of these studies researchers used a 7-item Likert type scale to compare pairs of artefacts participants built based on their similarity. The lower values on the 7-item Likert type scale in the studies were 1 and 0 which both indicated that the pairs of images were least similar. Whereas the high values, 7 and 6, indicated high similarity within pairs of images. In these studies, each pair was rated twice and checked for inter-rater reliability. The ratings were significantly correlated. However, it is important to add that correlations were weak probably due to the fact that similarity is hard to measure objectively (Caldwell and Millen 2008b). Both previous studies, that used a rating method, were interested in the similarity of structures between the same and different transmission chains whereas I was interested in the similarity score as a proxy for the originality of the builds within and between different experimental conditions. Therefore, I used a different type of 7-item Likert scale where 1 corresponded to the most similar pairs and 7 to the pairs being very different.

In order to collect data from raters through the online platform, Prolific, I developed an application in javascript using the jsPsych library (www.jspsych.org; de Leeuw 2015). I developed a web application since the number of pairs in all studies was high (> 10000 possible combinations). With the use of the web application I was able to automate the data collection process. The application enabled me to either randomly generate pairs of images for comparison (Study 1 and 3) or preselect the pairs (Study 2). First, the application randomly generated pairs from the pool of images. The same image was in some cases presented on the left of the screen and in some cases on the right. Second, randomly generated pairs were presented in a random order for each rater to compare. This enabled me to rate a large number of pairs of images in order to generate thousands of comparisons. Such an approach to increasing objectivity in the rating of similarity (or originality) would have been difficult to accomplish without the digital tool.

The application included an consent form (Appendix 4), a few questions about the rater (age, sex and Prolific ID) and the ratings task (Figure 8). The code for the application which I used to gather the similarity ratings of the pairs of structures can be found at https://github.com/piskotk/rating-creativity.

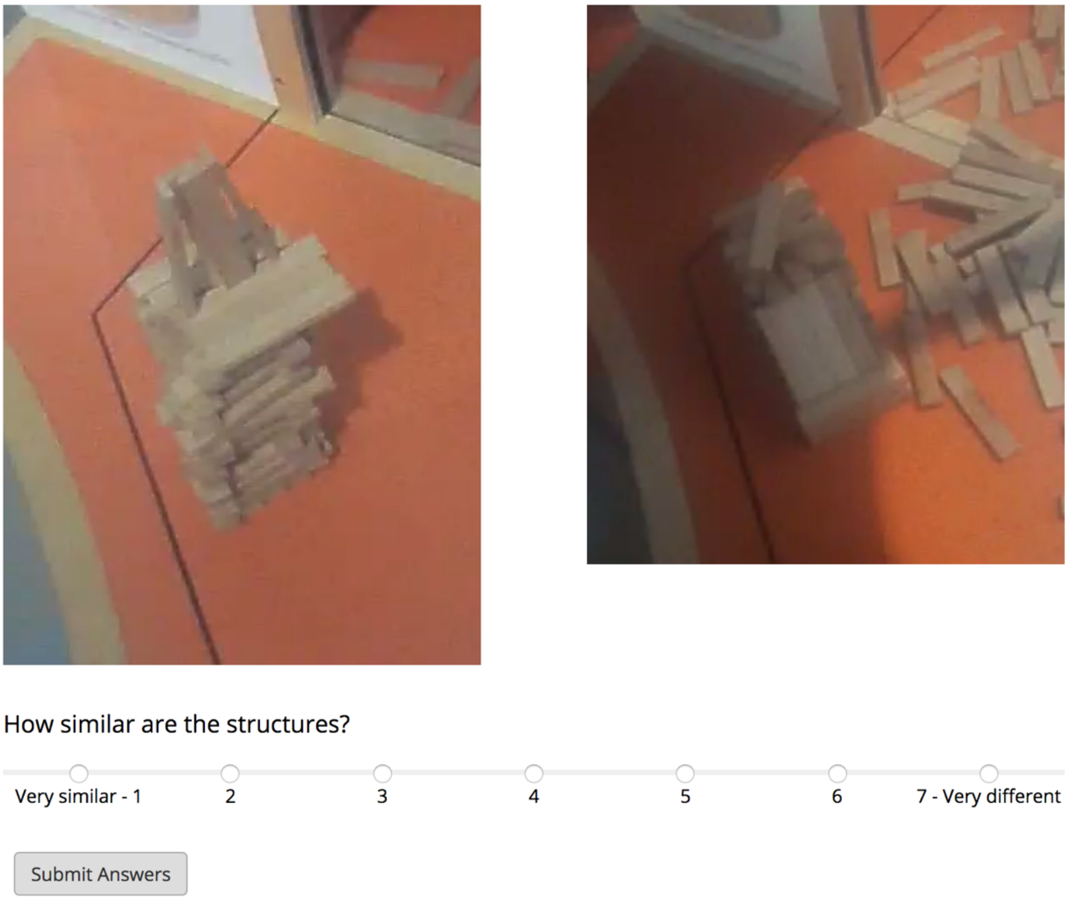


Figure 8: Screen shot of the task in which adult raters were asked to compare pairs of builds based on their similarity.

I used a similar type of measurement across my studies (Chapters 3, 4, 5) to establish a measurement of originality of pairs of structures children built in different conditions. In addition, in the first study (Chapter 3), I used a similar methodology to that described above to quantify the extent to which what children said they built (e.g. house) corresponded with the structure they built. I asked students at Durham University to rate individual images or pairs presented on a PowerPoint presentation. The raters were asked "Does it look like a house?" and filled out their answers which could be between 1 (no, not at all) and 7 (yes definitely) on a worksheet.

2.2.4.1 *Training set for raters and burn-in period from the ratings data collected through the web application*

I collected data on the correspondence and the ratings of similarity within the structures one individual built in Chapter 3. All raters received a training set of 10 example slides which I did not include in the analyses to ensure they familiarized themselves with the task and the structures they had to rate.

In the web application I did not provide a training set and raters only saw one example slide with instructions after which the task followed. To check whether there was a burn-in period, in which the responses of the raters plateaued, I performed an additional analysis with all sets of data. From these data sets I randomly assembled smaller data sets with at least 5000 responses in which I deleted either the first 10, 20 and 30 ratings provided by raters. I modelled the responses with an intercept only model and compared the models among themselves. The ratings that images received were not impacted by their position in the set and therefore, I used all the data in the analyses. The graphs of estimates from the models can be found in the appendix 6.

2.3 Ethical Statement

The Anthropology Department's Ethics and Data Protection Subcommittee awarded ethical permission to all studies. All studies included videos of children which were saved using numerical codes for each participant. The researchers who conducted the studies did not know the children's full names. The demographic data we collected were sex, age and the number of siblings. All data were safely stored on external discs that were locked in the office when not in use. The data that I collected automatically through the Interactive Research Pod were saved on a repository on a server at the LSC that had the latest firmware and was not connected to any other network but the internal system established within the exhibit. All consent forms can be found in the appendices (1, 2, 3) and all followed the guidelines for remote consent created by the British Psychological Society.

The responses from university students only included the ratings they assigned to different images. The data collected through Prolific included only the age and sex of the participant together with their Prolific ID and their answers to questions about builds' similarities. Each participant who started rating the images was first asked to consent to the collection and to their data being saved (the online consent form can be found in appendix 4). The online application only saved data from participants that agreed at the end of the task that their data could be used for research purposes.

2.4 Inter-coder reliability

Throughout the thesis 25% of data was second coded by a coder unaware of the conditions and the predictions of each chapter. I chose to second code 25% of the data since this is a practice used in most of the literature included in this thesis. First, I had practical constrains, since the lengths of the videos in Studies 1 and 2 (Chapter 3 and 4) ranged between 5 to 20 minutes for 340 participants in all three studies which would be impractical for a second coder to code all videos. Second, the nature of the data, natural observations in comparison to puzzle box problem solving, demands more attention and takes a longer time to code. I used the package irr (Gamer, Lemon, and Singh 2019) in R to compute the Cohen's kappa coefficient.

2.5 Bayesian statistical analysis

In recent years science has been full with debates about reproducibility and analyses (Open Science Collaboration 2015). One of the answers to the crisis is the development and implementation of alternative analytical methods which are not yet commonly used. Since newer computers and their power enable us to run more complex calculations (Witmer 2017), Bayesian analysis has come into the spotlight in psychology (Fullerton 2009). Bayesian analysis is not yet commonly used in evolutionary anthropology and in developmental psychology. However, there have been pleas to start using Bayesian analysis in research more broadly. This resulted in papers being published that present Bayesian analysis to newcomers (Kruschke and Liddell 2018), with annotated reading list suggestions (Etz et al. 2018) and articles specifically addressed to developmental researchers (van de Schoot et al. 2014). The value of Bayesian approach is that it infers distributions for the relationships between parameters of interest (predictor variables) and response variables. General Linear Models (GLM) approaches only focus on the mean and its confidence intervals, which are based on repeated sampling theory (for example a 95% confidence interval shows that out of 100 repetitions of the experiment, in 95 experiments the same results will be achieved). In contrast, Bayesian analysis provides an zone of posterior probability named credible interval, which can be interpreted as the probability that the parameter of interest is within the specific interval (van de Schoot and Depaoli 2014). Furthermore, the interpretation of the results in Bayesian analysis is not based on null hypothesis testing using p-values, but delivers the probability of a null hypothesis being true, which makes the interpretation of results easier.

In this thesis I built Bayesian statistical models to analyse my data. In Bayesian statistics the posterior distributions of a set of parameters of interest are calculated after considering the priors researchers set up for a set of parameters and the likelihoods the models receive from the set of parameters. The parameter(s) in the model is the unknown entity which we want to predict from the information contained in the data collected.

The prior describes the distribution of the parameter before we have the data and presents the prior plausibility assigned to each possible value of the parameter included in the model (McElreath 2016). Throughout the thesis I used weakly informed priors (Table 3) meaning that the distribution of the parameters of interest were not constrained by the model I used. This was appropriate because I had large data sets and I could not provide more informative priors as there were no previous studies similar to the ones presented in this thesis. The models and posterior distributions of the parameters were therefore heavily informed by the large data sets.

Table 3: Weakly informative priors used in all studies in this thesis

parameter type	prior (M, SD)
binary categories (e.g. sex)	normal distribution (0, 1)
continuous (e.g. age in months)	normal distribution (0, 1)
categories with more than 2 categories (e.g. condition, individual ID)	hyperprior[1] normal distribution (α, σ) α - normal distribution (0, 1) σ - cauchy distribution (0, 4)

[1] Hyperprior presents a prior distribution for the variation (normal distribution (α, σ)) in the prior distribution of intercepts α and σ. For example, α presents a prior for an average condition and σ presents a prior for standard deviation of conditions.

Bayesian inference uses priors and collected data to compute posterior probability which is a revised probability of a parameter having a specific response after considering new information (the collected data). This is done through an approximation process which estimates the distribution of the posterior probability. In this thesis I used approximation through Monte Carlo Markov Chains (MCMC) because some of the models I used were multilevel and included a lot of parameters (McElreath 2016). Under these circumstances, grid and quadratic approximation usually take too long to compute or do not work properly with more complex models. I ran models with 3 chains and 3 cores, the warm up period was set up to 1500 samples and the approximation was made with a subsequent 3500 iterations of real samples which were used for inference. Posterior distributions are presented either using a graph or through the highest posterior density intervals. All analyses for this thesis were performed in R 3.3.2 (R Core Team 2013) using either the *rethinking* package (McElreath 2016) or the *brms* package (Bürkner 2017a, 2017b). The *rethinking* package offers tools ('map2stan' function) to run Monte Carlo Markov chain (MCMC) through RStan 2.17.3 (Stan Development Team 2018).

I checked Rhat and the number of effective samples (n_eff) of all models. R_hat and n_eff indicate whether the models successfully converged (McElreath 2016) and could therefore be considered valid. In this thesis all models had effective data samples higher than 200 which presents the lowest limit from which a good estimate of posterior distribution could be taken. In cases when Rhat was more than 1 and n_eff was low, I checked trace plots of Markov chains to see whether they were clean, healthy Markov chains or not. In cases when the Markov chains were not clean (McElreath 2016, p. 258)

I either adjusted the delta and the tree depth or adjusted the priors in order to improve the model's convergence.

To evaluate and select the models, I used Akaike weights and WAIC values. The Akaike weight of the model is "an estimate of the probability that the model will make the best predictions on new data, conditional on the set of models considered" (McElreath 2016, 199). The values of the weights are spread between 0 and 1 and their sum is never higher than 1. The best possible model would have a weight higher than 0.9 which would suggest the model makes the best predictions of the new data. In cases when all of the Akaike weights in the model comparison were lower than 0.9, there is not a clear best model. The most effective model is probably a combination or ensemble of the effects of the models being compared, weighted by their relative Akaike weights. However, when fitting models to the data one needs to be conscious of the research question. I calculated and displayed the posterior distributions from the model which was most suitable for the specific research question. Widely Applicable Information Criterion (WAIC) estimated out-of-sample deviance, the smaller the value the better the model was in predicting the parameter of interest (McElreath 2016).

I presented results through three types of graphs: posterior distribution (Figure 9), probability of a specific outcome (Figures 10, 11) and marginal effects graphs (Figure 12).

2.5.1 Posterior distribution graph

The graph (Figure 9) is taken from Chapter 3 where I explored the effect of direct instructions on children's exploration and manipulation of blocks and the originality of the structures children built. I use it here to illustrate and present the type of graph which is used in this thesis. The graph presents a posterior density plot showing the probability of an outcome (in this case manipulating the blocks) on the x-axis with their relative likelihood represented by the density on the y-axis. The graph indicates that it was most likely for participants not to perform a manipulation of a block when in the condition with no instructions, whereas it was least likely they did not manipulate blocks in the condition with instructions. This is seen by the posterior distribution being shifted to the left for instructions and to the right for the no instruction condition. The scaffolding condition, in which I guided children to explore the block's affordances (the fact that the blocks can change shape), shows that the children were equally likely to manipulate as to not manipulate blocks (Figure 9). The distribution is fairly narrow and thus can be interpreted as evidence supporting ambivalence on the children's part, as opposed to uncertainty over the probability of not doing any block manipulations (which would be concluded if the mean of the distribution approximated 0.5 but with a broad distribution).

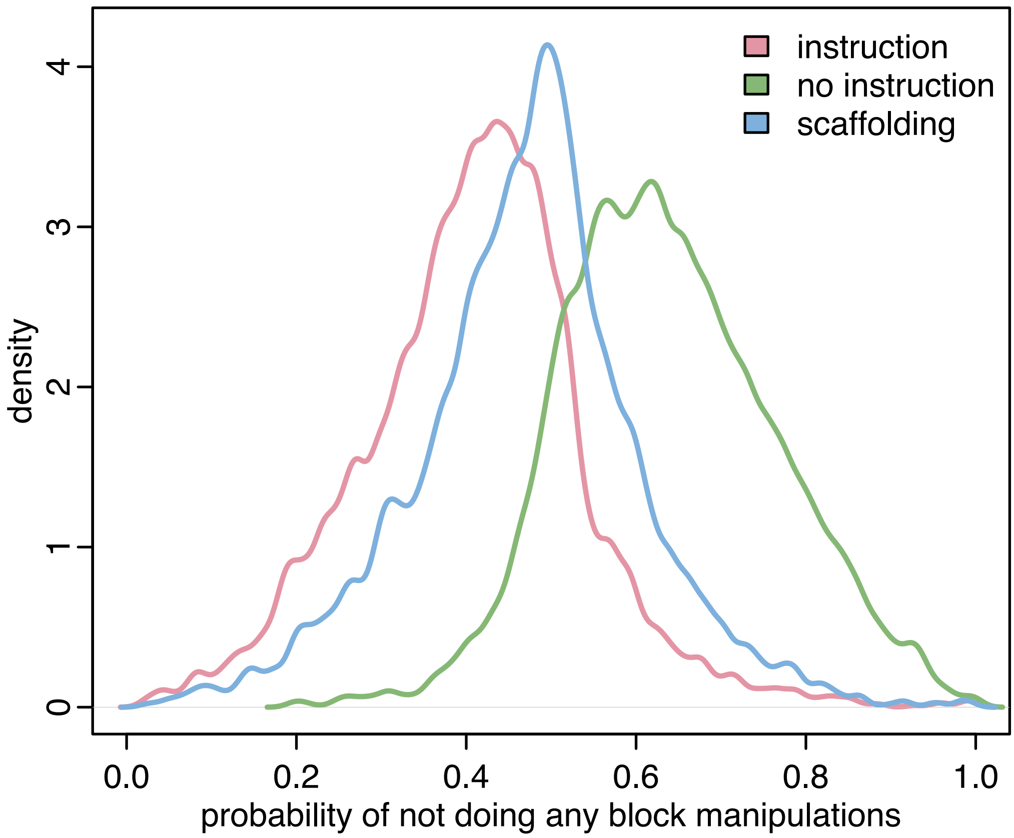


Figure 9: Posterior density plot of the probability of not doing any manipulations of blocks in different conditions.

2.5.2 Ordered categorical regression

A Likert scale is a type of measurement common to research that uses questionnaires. Typically, social science researchers are interested in their subjects and their responses to survey questions. In this study I used a Likert type scale to assess whether structures children built out of wooden blocks resembled what they stated they had built and to assess the similarities between the structures. Likert type scales usually offer responses on a scale, for example 1 to 5. In the studies presented in this thesis I used a 7 item Likert type scale. When individuals answer surveys on Likert type scales they do not understand responses as having the same distance between two values, but as categorical and not necessary normally distributed (Fullerton 2009). Therefore, we need to perform an ordered categorical regression when conducting the analysis (Liddell and Kruschke 2018; Bürkner and Vuorre 2018).

In the ordered categorical regression the response variable is the Likert item from 1 to 7 and the predictor variables are the child's age sex and the condition etc. The basic intercepts are the thresholds for each of the states of the order (of 1-7). Following McElreath (2016), Bürkner and Vuorre (2018), I used a cumulative approach to the comparison of the response variables, which compares the probability of being at or below a certain rating to the probability of being beyond that rating. There are 7 minus 1 (6) intercepts (thresholds that need to be overcome to get to a higher rating). The ordered categorical regression is composed of several (the number of categories of Likert items minus 1) binary logit functions and each threshold presents a comparison of the probability being less than or equal to the threshold (in comparison to the probability of being beyond that rating). To present results from this analysis I used graphs showing the posterior predictions of ordered categorical models (Figure 10) and marginal effects graphs (Figure 11).

2.5.2.1 Posterior predictions of ordered categorical model graph from the rethinking package

The graph showing the posterior predictions presents the probability of an outcome on a 7-item Likert type scale (Figure 10). Figure 10 shows how the predicted response (the value on the Likert type scale) varies with a change in a predictor variable, parameter X. The model predicted 7 response values on an ordered categorical scale. The grey lines show samples from the posterior distribution for the cumulative probability of reaching each threshold between level (from 1 to 7) on the Likert type scale. We see how the thresholds change with parameter X. As the vertical axis shows the cumulative probability, the probability of a specific response value being selected is presented in between two adjacent clusters of lines, so the probability of a Likert response of level 1 is given by the lowest cluster of lines, while the probability of a Likert response of level 2 is the vertical distance between the bottom cluster of lines and the adjacent cluster of lines. The vertical distance above the top cluster of lines indicates the probability of a Likert response of level 7. If parameter X has a binary value, then the graph shows that the probability of the level 7 on the Likert scale being selected is higher when parameter X is 0 than when it is 1. In this toy example, the threshold posterior distributions are separate from one another, but in 'real life' they can overlap if the posterior distributions for each threshold are broad: as the distributions show the cumulative probability, the rule of thumb is that if the lines are low down on the vertical axis, the posterior distribution is predicting *high* Likert levels (i.e. most of the probability is above the threshold between levels 6 and 7), while if all the lines are stacked up at the top of the graph, most of the probability is associated with the *lowest* Likert level. An even distribution of line clusters indicates an approximate even distribution of Likert level responses.

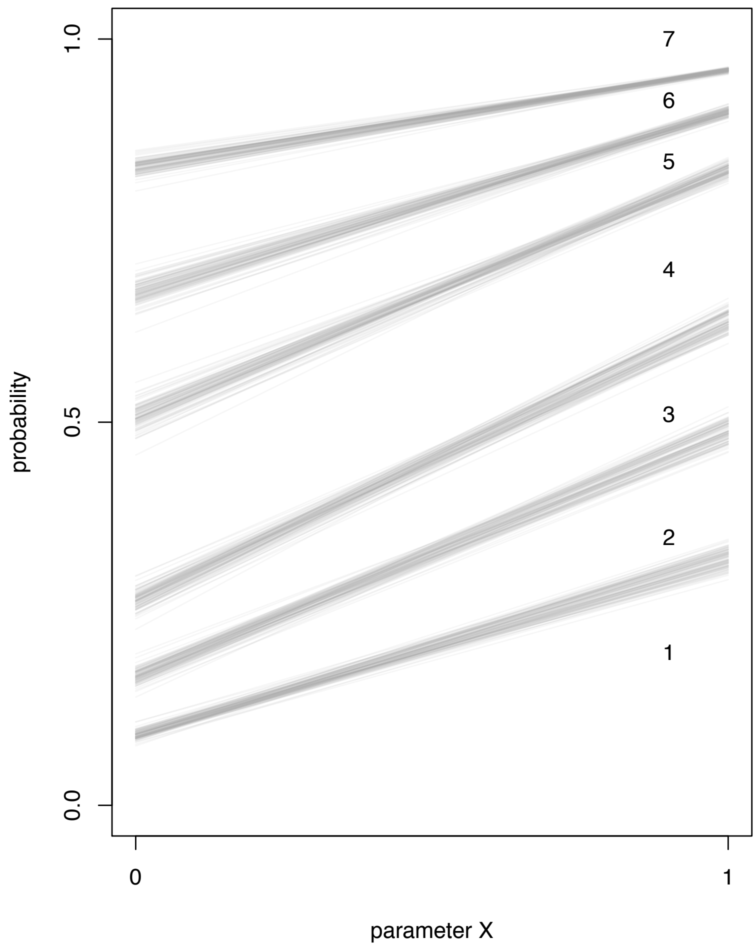


Figure 10: Posterior prediction of the ordered categorical model with the predictor parameter X. The lines represent the boundaries between response values, numbered 1 to 7. This example is taken from Statistical Rethinking (McElreath, 2016, p. 342).

For clarification, I am adding an example of the graph presented in the Chapter 4 (Figure 11). The vertical axis represents the cumulative probability of ordinal categories of the outcome variable. In the studies presented in this thesis I used 7 categories to describe the similarity of the structures children built. The cumulative probability of these categories is spread between 0 and 1 on the vertical axis. The analysis shows allocation of the cumulative probability among the categories. Greater cumulative probability might be taken at lower ordinal categories (higher similarity) or higher categories (higher originality). Grey lines show where predicted thresholds between the ordinal categories lie in terms of cumulative probability. Distributions show cumulative probability, when they are stacked more towards the top of the graph this shows the structures compared were more likely to be rated low on the Likert scale (as very similar).When lines are stacked more towards the bottom, the cumulative probability is bigger for higher values on the Likert scale (higher originality between structures compared).

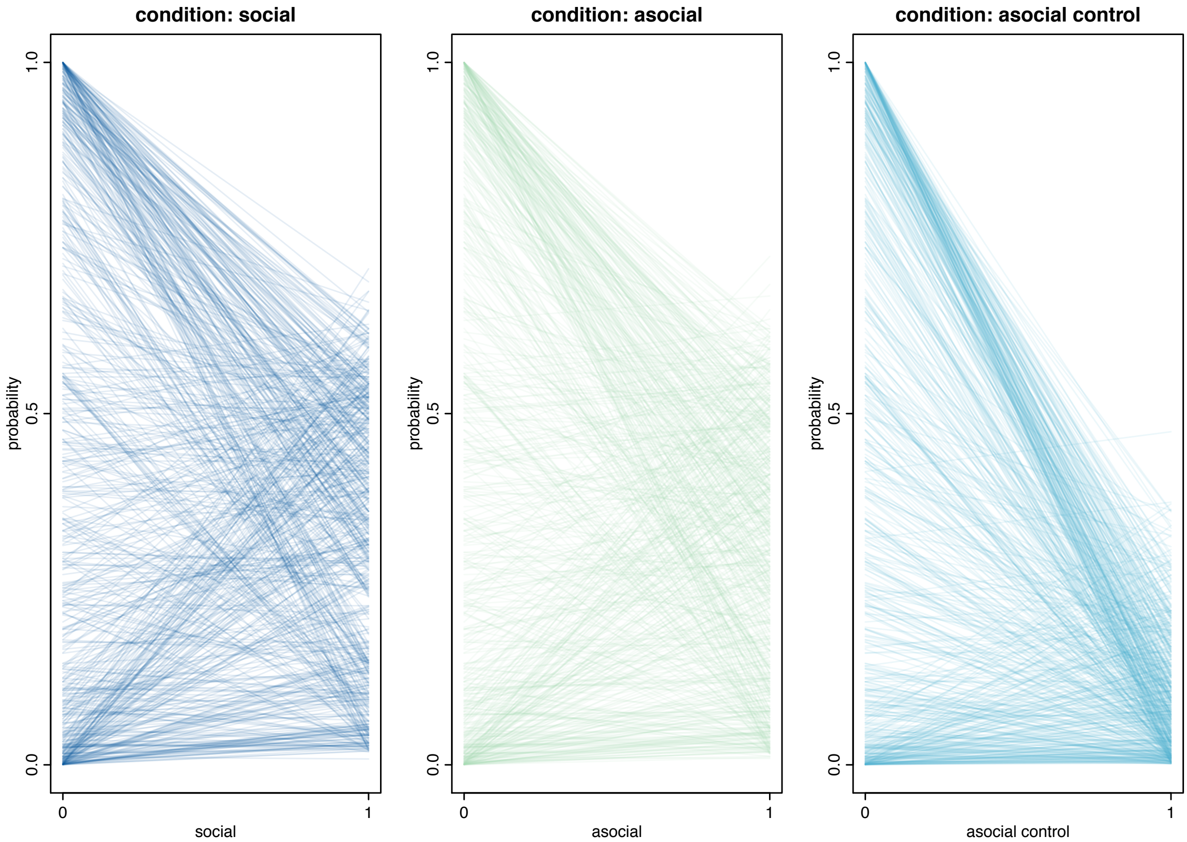


Figure 11: Posterior prediction of the ordered categorical model with the predictor parameter as condition. The lines represent where predicted thresholds between the ordinal categories lie in terms of cumulative probability. Distributions show cumulative probability, when they are stacked more to the top of the graph, this shows the structures compared were more likely to be rated low on the Likert scale (as very similar), whereas when lines are stacked more to the bottom, the cumulative probability is larger for higher values on the Likert scale (higher originality between structures compared).

2.5.2.2 Marginal effects graph from the brms package

The marginal effect plot of the ordered categorical model shows the probability of the rating being selected by the individual (in this thesis the rater) under specific conditions. This example is taken from Chapter 3, where rates provided ratings of the similarity of the structures built by the same individual. Probability in the example (Figure 12) is divided between seven categories that correspond to the ratings on the Likert scale.

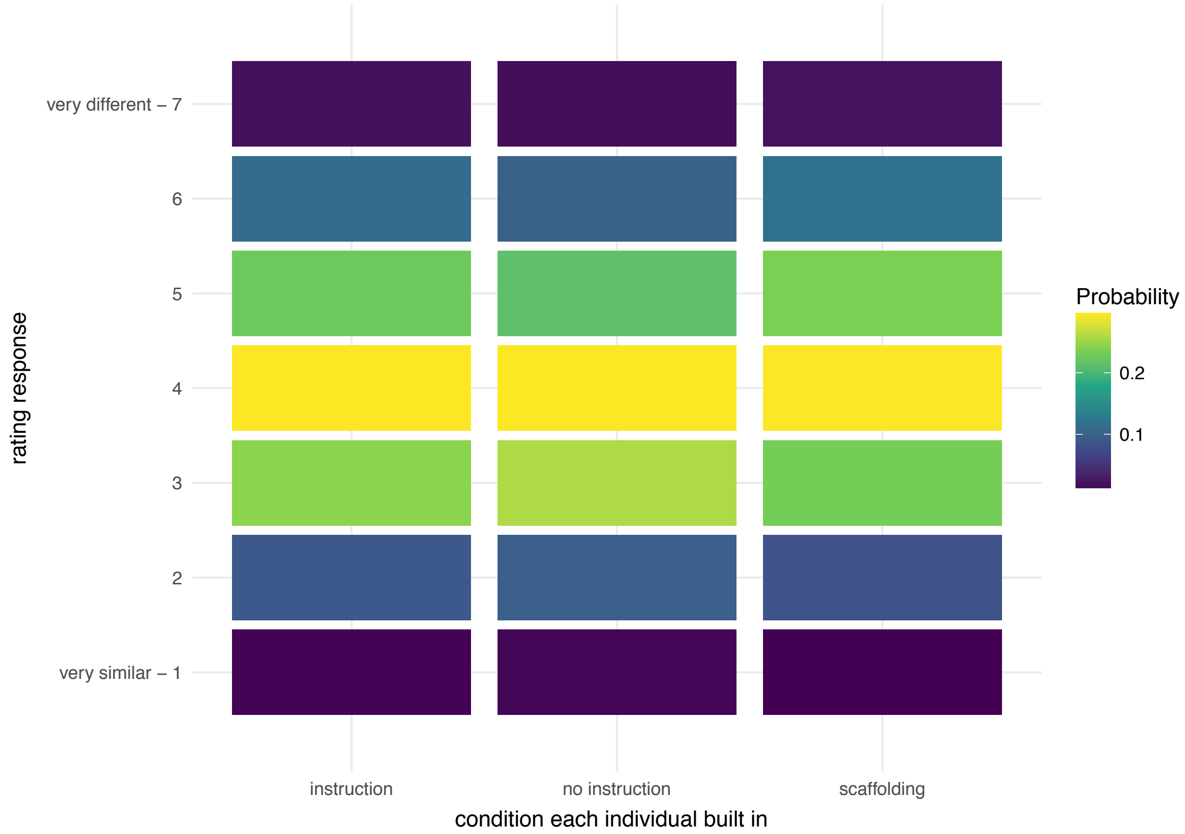


Figure 12: Marginal effect plot from *brms* package. Different colours correspond to different probability of specific rating being selected by the raters.

3 Do instructions squash creativity?

Children's creativity is not susceptible to instruction and scaffolding

Abstract

Children may engage in different types of learning in informal learning environments. It is known that children are good imitators and explorers. They might also receive guidance from science explainers, their guardians or peers. How these types of guidance affect their creativity is not known. I predicted direct instructions would limit creativity, while scaffolding type instructions and no instructions would foster it.

In this study 172 children aged between 4 to 12 years either experienced direct instruction on how to use wooden shape shifting blocks, scaffolding (with open questions) or no instruction at all. After that they were asked to build three structures (whatever they wanted) from wooden blocks. I collected ratings of the similarity between structures children built and analysed whether there were any connections between the condition they were in and originality of their structures.

The condition (type of instructions or no instructions) did not affect the originality of the final structures built when comparing within individuals, between individuals in the same condition or between individuals in different conditions. Given the literature, this was unexpected and may have been due to the vagueness of the instructions employed or the fact that the task was not goal-oriented and easy compared to tasks used in the literature.

3.1 Introduction

As human beings, young children are curious, they explore the world that surrounds them, solve problems, test hypotheses and are active learners (Bonawitz et al. 2012; Gopnik 2012). However, they do not seem to be remarkably successful tool innovators and problem solvers (Beck et al. 2011; Nielsen 2013; Hanus et al. 2011; Tennie, Call, and Tomasello 2009; but see Reindl and Tennie 2018). Innovation can be defined as something that is new, useful and will potentially spread through group as a learned behaviour (Carr, Kendal, and Flynn 2016). Many processes may underlie innovation (for a review see Carr, Kendal, and Flynn 2016), two of them being creativity and divergent thinking. There are many definitions of creativity some of which focus on novelty, appropriateness and broader impact (e. g. influencing a group of individuals and not just one) (Piffer 2012; Diedrich et al. 2015). The most commonly used definition is that creativity is something that is novel either to an individual or to a population (Carruthers 2002). Behaviours are more likely to be novel to children than adults, simply because children are exposed to fewer behaviours and experiences than adults. In addition, divergent thinking is the ability to generate numerous diverse ideas for a single problem (Guilford 1959) and as such, may influence creativity. For the purpose of this chapter I define creativity as something that is novel.

An important domain of creativity, exploration, was investigated in children using different puzzle boxes (Bijvoet-van den Berg and Hoicka 2014; Bonawitz et al. 2011). When children, as young as two years old, were given unfamiliar objects, they would explore them and even undertake divergent actions (Bijvoet-van den Berg and Hoicka 2014). This supports the premise that children do in in fact possess abilities for divergent thinking at a very early age (Hoicka et al. 2016). However, divergent thinking is not necessarily a reliable predictor of the creativity of an individual (Runco and Acar 2012; Beck et al. 2016). Children also express creativity through collaboration and invention of their own social norms, for example determining the way the game, marble run, 'should' be played (Göckeritz, Schmidt, and Tomasello 2014). This also highlights that children might be more engaged with tasks that have meaning to them.

A large part of childhood is being guided and instructed by family members, friends, peers and teachers. Several studies indicate that the use of instructions or demonstrations delivered by adults, especially when children explore novel objects, limits their explorative behaviour (Bonawitz et al. 2011; Shneidman et al. 2016; Wood, Kendal, and Flynn 2013a; Wood et al. 2016; but see Butler and Markman 2012). A Bayesian model of pedagogy correctly predicted the extent to which pedagogy limited 4 to 6 year old children's exploration and innovation (Bonawitz et al. 2011). When children were not presented with any instructions, they explored a novel toy for longer and discovered more of its functions than children that received instructions. Children were inhibited by instructions even when they were not delivered directly to them, but to another child in the room (Bonawitz et al. 2011). Moreover, in a different study, children who were exposed to one demonstrated method of solving a novel puzzle box task were more likely to use that one method of solving the task and did not explore the box further to discover other methods compared to children who were not exposed to demonstrations (Wood, Kendal, and Flynn 2013a). Likewise, children who were exposed to demonstrations that comprised of irrelevant causal actions, instead of direct instructions, also decreased their exploration of alternative solutions to the problem (Wood et al. 2016). However, even if children explore less when given instructions or demonstrations, sometimes such instructions or guidance might provide them the opportunity (or scaffolding) to develop and create novel things, as was shown with toddlers making up jokes (Hoicka and Akhtar 2011).

In children older than 6 years research regarding the impact of instructions on their learning is conflicting and research about whether instructions influence innovation is lacking. Children may actually develop formal rigorous scientific thinking with the help of instructions and guidance (Kuhn 1989). For example, when children are challenged to control the variables in a mock scientific experiment, 9 to 10 year olds that received instructions fostered a better understanding of the experiments than did those who could only employ discovery learning (Klahr and Nigam 2004). Conversely, a longitudinal study with children of the same age showed the benefits of individual learning through discovery (Dean and Kuhn 2007). Time scales may be important here, however, as the children had a few months to learn in the latter study, resulting in discovery learning of scientific concepts being as effective as learning with instructions.

Studies with children within informal learning environments, such as science centres, have shown conflicting results. When it comes to interactions with exhibits, in some cases instructions limit exploration and active manipulation while in others, they promote it. In a science centre in Amsterdam (NEMO), a comparative study was carried out of two exhibits with which children aged between 4 to 6 years interacted in three different conditions (Van Schijndel, Franse, and Raijmakers 2010). Children's parents were asked to provide their children either with instructions, or to guide them through the exhibit with open questions and directing their attention (scaffolding) or with minimal instructions. The researchers compared exploratory interactions with two exhibits, these were the 'Rolling, rolling, rolling' and the 'Spinning forces'. The first consisted of a wooden deck with three descending tracks that are covered either with artificial grass, carpet or nothing. Visitors were provided with three different cylinders and were prompted to explore the effect of the deck covering and the type of cylinder on how they rolled down the tracks. The latter, the 'Spinning forces' exhibit, comprised of a chair and two wooden blocks. The chair could spin around and visitors explored the effect of the body and block position on the speed of spinning. The researchers developed an exploratory behaviour scale based on the children's interactions with the exhibit. Children could either adopt passive or active contact with the exhibit or actively explore it. The results differed between the exhibits and the conditions the children were in. Children explored most when minimally guided with the 'Rolling, rolling, rolling' exhibit but explored the 'Spinning forces' more when provided with full instructions. The intermediate scaffolding condition promoted the active manipulation of the exhibit but did not lead to exploration. In a similar study at the NEMO science centre, children's exploration of shadow sizes in the "Shadow" exhibition was supported when parents provided children with descriptions of evidence, which included exhibit features and their observations (for example: "that shadow is smaller than the other shadow") (Van Schijndel and Raijmakers 2016). Whereas, providing causal explanations or instructions that would direct children as to how to use the exhibit (for example if you put the car closer to the light, the shadow gets bigger) did not foster exploration (Van Schijndel and Raijmakers 2016). Evidence descriptions are similar to showing children affordances of objects or tools which besides promoting exploration also facilitates innovation (Neldner, Mushin, and Nielsen 2017). In a different study, when children investigated an interactive exhibit together with their parents, they explored them for longer and more broadly than children who explored exhibits on their own or with peers (Crowley et al. 2001). Here, the type of exhibit may have been influential as it had a clear aim and there were specific ways that children could explore it and therefore, parental involvement supported the interaction. Likewise, in a task where children had to use diagonal blocks to strengthen a tower, instructions presented as analogies positively impacted problem solving abilities in children (Gentner et al. 2016). All of these tasks had a clear goal, but science museums also include open-ended hands-on exhibits.

Open-ended science exhibits can be freely explored by the visitors and do not have a limited number of solutions nor any instructions on how to use them. It is known that open-endedness, together with the technological novelty of an exhibit, holds visitor's attention for longer than exhibits that are user-centred and have a lot of sensory stimulation (Sandifer 2003). Spending more time at an exhibit might promote explorative and creative behaviours. The impact of instructions and guidance on children's exploration and creativity when interacting with open-ended exhibits has not yet been studied. There has, however, been limited research in adults. Research with adults indicates that creativity is constrained when participants receive instructions about what to build with LEGO bricks (Moreau and Engeset 2015). An open-ended task with bricks, on the other hand, fostered creativity in adult participants (Moreau and Engeset 2015). Besides using clear instructions, science centres also introduce some exhibits through the use of signs that provide scaffolding and open-questions. There is little evidence of how open questions impact visitor's interactions with exhibits and whether they influence children's exploration and creativity. This is a pertinent question theoretically and practically as science centres that provide open-ended exhibits are often criticised by parents for not providing 'instructions' (Andy Lloyd pers. comm.).

Most previous studies investigating children's creativity and innovation used rigorous experimental setups (Carr, Kendal, and Flynn 2015; Beck et al. 2011; Nielsen 2013; Tennie, Call, and Tomasello 2009). Children were mainly tested on tasks with clear goals and limited solutions in schools and in unfamiliar settings of research institutions (Carr, Kendal, and Flynn 2015; Beck et al. 2011). Here I used a more ecologically valid means of investigating children's creativity, and what influences it, through an open-ended task within an informal learning environment (a science centre in the NE of England, Life Science Centre). Instead of using a task with a clear goal, I used an open-ended task based on an exhibit containing wooden 'shapeshifting' blocks. The children received either (i) direct instructions regarding the affordances of shapeshifting blocks, (ii) open questions to prompt exploration of affordances (scaffolding), or (iii) no instructions at all.

Since there are mixed findings in the literature regarding the impact instructions have on children's creativity, I made tentative predictions based on the fact that the task I presented children with did not require specific prior knowledge.

Direct instructions:

> I predicted that providing children with direct instructions regarding how to use the blocks would support their active manipulation of blocks, but would also limit their exploration of them (Bonawitz et al. 2011; Wood, Kendal, and Flynn 2013a).

I would expect instructions would facilitate originality of the structures, because children would be familiar with the affordances of blocks (Neldner, Mushin, and Nielsen 2017).

Scaffolding (Open questions):

I predicted that the scaffolding condition would promote active manipulation of the blocks and also exploration of the blocks (Van Schijndel, Franse, and Raijmakers 2010).

Due to the inconsistencies in findings across fields and since I used an open-ended task I did not make specific predictions regarding the relative extent of originality in scaffolding condition.

No instructions:

I predicted that in the condition with no instructions children would perform less directed manipulation of the blocks (because the variety of affordances are not obvious), but this would not limit their exploration of the blocks (Bonawitz et al. 2011).

Due to the inconsistencies in findings across fields and since I used an open-ended task I did not make specific predictions regarding the relative extent of originality in no instruction condition.

Divergent thinking and originality of the structures individual built:

Finally, I investigated the relationship between the originality of the structures and divergent thinking in children. I predicted that the measures of divergent thinking would not be predictive of the final structure originality scores since divergent thinking does not predict the success of children between 4 to 8 years old in problem solving tasks (Bijvoet-van den Berg and Hoicka 2014) and divergent thinking cannot be understood as the same as creativity (Runco and Acar 2012).

3.2 Methods

3.2.1 Participants

One hundred and seventy-four children (aged 4 to 11 years; 91 males) were recruited within Life Science Centre during weekends and school holidays between November 2015 and February 2016 (Table 4). Three children were excluded from the analysis since they were not able to interact with and change shapes of the blocks and two children were excluded because of issues with video recordings. The study was conducted in a section of the science centre that contained open-ended and hands-on exhibits (see section 2.1.2 for a detailed description). Consent was provided by the children's guardians (Appendix 1) and partitions were used to provide an experimental space with as little distraction as possible (Figure 2).

Table 4: The profile of participants

Age	No instruction	Scaffolding	Direct instruction
4	6 (3F)	7 (3F)	7 (4F)
5	9 (3F)	10 (2F)	7 (2F)
6	14 (4F)	12 (8F)	13 (8F)
7	10 (5F)	14 (8F)	14 (7F)
8	8 (5F)	10 (6F)	8 (5F)
9	4 (2F)	3 (1F)	5 (2F)
10	6(3F)	3 (1F)	2 (1F)
11	1 (0F)	0	1 (0F)
Total	58 (25F)	59 (29F)	57 (29F)

3.2.2 Apparatus

3.2.2.1 Divergent thinking

To test divergent thinking, I used Guilford's alternative uses task (Guilford 1959). In an introduction phase I presented the child with a white rope (length = 20 cm) and asked her/him to tell me as many alternate ways of using it. Once understanding was established, the testing phase began and I asked the child how many different ways s/he could use a plastic cup (volume = 2 dcl). Data were collected from the second test only. I counted and scored the possible uses given for four characteristics: originality, fluency, flexibility and elaboration (Table 5).

Table 5: Explanation of the four characteristics scored in divergent thinking tasks

fluency	sum of all responses
flexibility	number of different categories of responses
originality	comparison of every response provided by a child to the total amount of responses children provided; responses given by 1% of the tested group counted as unique response and were given 2 points, responses given by less than 5% of the group tested were given 1 point; higher scores indicated creativity; I corrected the originality score by dividing originality with fluency. Since if left uncorrected, the fluency and originality positively influence each other.
elaboration	amount of detail of the use for a cup and/or subject with wheels

I used Wallach and Kogan's test (1965) as the second measurement of divergent thinking. I asked the child in an introduction stage to name things that could fly. In the testing phase I first asked her/him to name different things that had wheels and second, name different things that made a noise. All responses were counted and scored in the same way as with Guildford's alternative uses task (Table 5).

3.2.2.2 Wooden 'shapeshifting' block building methods

The novel building block task was based on an existing exhibit in the Life Science Centre. The exhibit comprised of wooden blocks consisting of smaller cubes connected via a rope through the centre (Figure 3). Children were able to change the shapes of the connecting blocks by placing cubes together in different ways (Figure 13). Blocks could be put together in different configurations to build different structures.

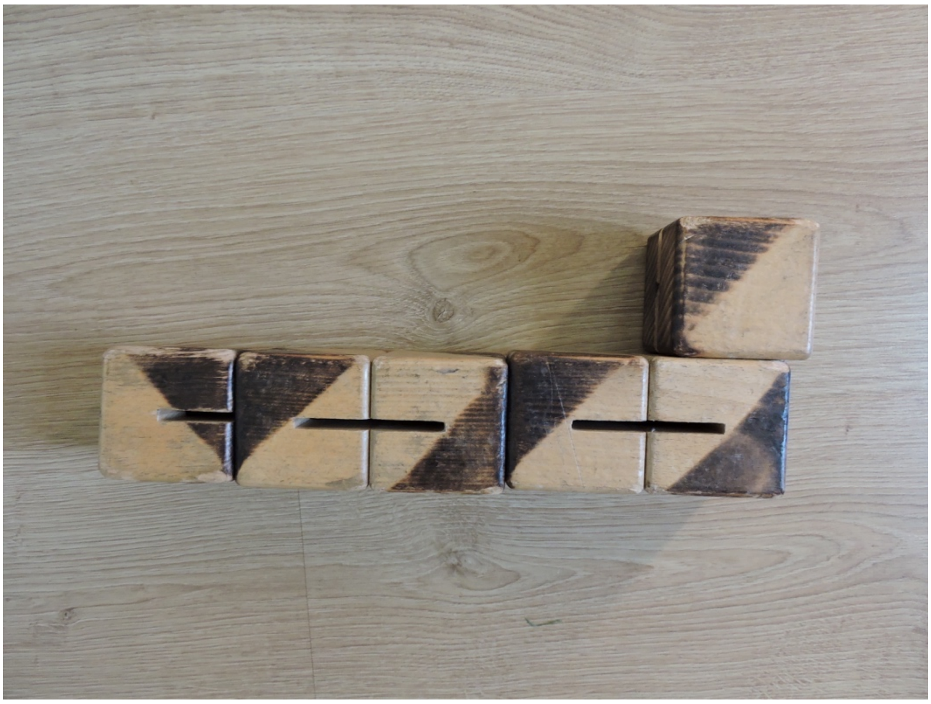


Figure 13: An example of a block in a changed configuration

3.2.3 Procedure

First, I introduced myself to the child and asked them questions about their age, previous visits to the science centre and whether they played games relevant to the task, such as video games (eg. Minecraft) or with LEGO bricks. After the introduction period children were given two types of tasks, the order of which were counterbalanced: (1) building using wooden blocks or (2) completing the divergent thinking tasks.

In the wooden block-building task, I introduced children to seven blocks of different lengths at three different building stations which were set one meter apart on the floor. I asked children to build whatever they wanted at each station and gave them a time limit of two minutes for each build. All children received the same introduction from the researcher. I asked them to build whatever they wanted and reassured them that they could not do anything wrong with the blocks. I asked them what they built after they had finished building at all three stations. Prior to this, the children were placed in three different conditions where they received different treatments. In the no instruction condition the children had an opportunity to touch and manipulate pairs of blocks to familiarize themselves with them. I said, "You can start now. Before we start, here are some blocks that you can look at and touch if you want." In the direct instruction condition, I demonstrated how to change the shapes of the two blocks, in a so-called shapeshifting action and how to put pairs of blocks together in a specific way. I said, "Before you start building, let me show you that you can change the shape of the blocks. Look and see the rope. You can put them together like this (visual presentation). You can start now." In the scaffolding condition I encouraged the participant to discover how the block's shapeshifting mechanism worked by asking, "Before you start building, can you work out different ways to change the shape of the blocks and put them together? Can you put them together in a different way? Give it a go", but did not physically show them anything.

After the child finished building I asked her/him “what did you build?” I then tested their ability to change the shape of the block (Table 6).

The children's behaviours and actions expressed while interacting with blocks were videotaped. Videos were coded using an ethogram of the children's behaviours (Table 6) and the frequency of these behaviours.

Table 6: Ethogram of observed children’s behaviours when completeing the building block task

Behaviour - Interaction with one block (separately for each pile)	
Twisting	participant holds a cube (a part of the block) and rotates it in the opposite direction to the other cube (part of the block)

Shapeshifting	changing the position of cubes that constitute a block; participant holds with one hand a part of the block under or over the part of the block the participant wants to move, and moves the cube of the block under or over it, the part of the block in between the hands changes position
All direct manipulations	sum of all direct manipulations (twisting and shapeshifting) expressed with one pile of blocks
Indirect manipulations	participant does not hold a part of the block, but a part of the block gets twisted passively; participant does not actively engage with the parts of the blocks that change shape, the parts of the blocks move by chance or participant used force on other part of the blocks; participant holds two parts of the block, but parts between the hands do not change shape, shape changes at the part where the participant is not actively holding the block; shape changes just by chance, when participant holds the whole block with just one hand, maybe the block fell on the floor
Explorative manipulations	when participant uses direct shapeshifting and twisting but does not use the changed shape of the block in the construction s/he is building; when they do not use it in the construction they are building, they just explore the affordances of the blocks; when they change the shape of the block and put it back to its original shape
Explorative fitting	when participant try to use block in construction, but before they put it into construction they adjust the shape again (e.g. they change shape, try to put it in the construction but it does not fit, they change shape again and it fits; the first part of action is explorative fitting)
Explorative actions	sum of explorative manipulations and explorative fitting
Test of ability to change shapes of the block	

Ability[2]	yes - child successfully repeats the action the experimenter made and moves part of the block so it changes its shape to shape L no - child in not able to move part of the block, the block does not change shape
Number of used blocks	
Number of blocks from the pile used to build construction	from 1 to 7
Number of straight blocks used in structure	from 1 to 7

Behavioural counts were converted into rates, to account for differences in the length of time that children spent building. The total number of different behaviours were calculated to evaluate how diverse the process of building was for each participant and in each condition.

3.2.3.1.1 Inter-coder reliability

To evaluate the reliability of the coder an additional 25% of the randomly chosen behavioural data were rated by another researcher unaware of the conditions and the predictions. These two coders rated the behaviours, all direct manipulations, exploratory behaviours, same and different manipulations. A Cohen's kappa of .957 was achieved, demonstrating almost perfect reliability.

As an additional test to classical Cohen's kappa coefficient following Bayesian statistics methods to evaluate the behavioural counts I compared WAIC values and their weights between an intercept only model and a model including coder ID as a predictor. All models used Poisson distribution, had weakly informative priors and were

[2] Three children were not able to change shapes of the blocks therefore I excluded them from the analyses.

approximated by a Gaussian distribution with 'map' function in *rethinking* package (McElreath 2016). There were no difference between the intercept only and coder ID models (Table 7). Therefore the coding was reliable.

Table 7: Inter-coder analysis, comparison between the intercept only and intercept and coder model.

behaviour	model	WAIC	pWAIC	dWAIC	weight	SE	dSE
manipulations	mIntercept	1719.8	5.6	0.0	0.94	75.84	N/A
	mCoder	1725.4	10.3	5.6	0.06	76.28	0.96
exploration	mIntercept	996.9	3.4	0.0	0.82	53.12	N/A
	mCoder	1000.0	5.7	3.1	0.18	53.54	1.36

3.2.4 Rating creativity

I developed a standardized and objective technique to assess the 'correspondence' (does it look like what child said s/he built) and originality of the children's final wooden block structures (Section 2.2.4).

3.2.4.1 Rating originality of pairs of structures

I recruited 56 undergraduate students in the Anthropology department (Durham University) who attended the course 'Our Place in Nature' to measure the originality of pairs of structures built by the same individual. I first introduced them to ten examples of the pairs of structures to practice on. These ratings were not used in the analysis. The ratings of the pairs of structures individual built enabled me to gain an approximation of each child's divergent thinking when asked to build with blocks.

To rate block structures between individuals within the same condition and between conditions, I decided to gather data only from the first, of the three, structures each child built as these were the most influenced by the condition the children were in.

I then recruited 232 (101 males; median age 31 years) participants through Prolific to measure the originality of pairs of structures within conditions and between conditions. If I compared 174 structures to each other, this generated 15051 pairs of structures to compare. To make sure each structure was rated at least a few times, I had to recruit a higher number of raters and automate the process. Therefore, I used Prolific and developed a web application written in javascript using jsPsych library (de Leeuw 2015) (Section 2.2.4). Each individual recruited through Prolific received a reimbursement of £1.25 for their participation. The web application randomly generated pairs of structures that were always from different individuals but either from a pool of structures made in the same condition or between conditions. I asked adult participants to rate pairs of structures on a scale from 1 (very similar) to 7 (very different) based on how structurally similar they appeared in comparison to each other (Figure 8).

3.2.4.2 Rating 'correspondence' between appearance and intended structure

I recruited 48 undergraduate and postgraduate students in the Anthropology department at Durham University who attended the course 'Our Place in Nature' to measure the correspondence between what the children reported they built and the degree to which this was discernible by the adult rater. Each adult participant was asked whether the structure in the picture looked like what the child had stated it was, on a scale from 1 (not at all) to 7 (yes definitely). Adult participants scored 522 structures each in this manner such that each structure was rated by at least six different adults.

3.2.5 Statistical analysis

To test predictions, I run Bayesian statistical models using the 'map2stan' function in the *rethinking* R package (McElreath 2016). Models were fit using Hamilton Markov Chain Monte Carlo in r-STAN 2.17.3 (Stan Development Team 2018) in R 3.3.2 (R Core Team 2013). I used the Watanabe-Akaike information criterion (WAIC) to compare the models using 'compare' function in rethinking.

Models of ratings were run using *brms* package (Bürkner 2017a, 2017b) in R.

Inter-coder reliability with Cohen's Kappa coefficient was calculated using the *irr* package (Gamer, Lemon, and Singh 2019) in R.

3.2.5.1 Behavioural variables

Behaviour from each individual was measured separately three times corresponding to the three different block structures s/he had to build. The predictor variables included the participant's age, sex (dummy variable), number of previous visits to the museum, what s/he did first either divergent thinking tasks or building blocks task and data pertaining to the divergent thinking task and block building task. The divergent thinking tasks included measurements of flexibility, elaboration, fluency and originality for each of the objects she was tested on. The block building task data included the condition participant was in (instruction, no instruction and scaffolding), the number of block manipulations s/he was exposed to (in instruction: 1, no instruction: 0) or s/he performed (in the scaffolding condition: max 2) before the actual building phase, time spent building in seconds, count of all direct manipulations of blocks, all explorations (sum of exploratory manipulations and fitting), number of blocks used in the final structure and number of straight blocks in final structure. I included individual ID and their structure (whether it was the first, second or third) as two separate categorical variables.

Table 8: Parameters included in models of behaviours

T	time building
I	individual
A	age (4 to 11 years old)
S	sex (0 - female, 1 - male)
B	first, second or third build (or structure ID)
F	what participant did first? (divergent thinking tasks or building blocks task)
Mv	how many times previously had visited the science centre? (first time, one to five times, more than five times)
C	condition (1 – instruction, 2 – no instruction, 3 – scaffolding)
DT	divergent thinking task scores

Age and time spent building were log transformed for the analysis. Log transformation of age enabled me to see whether older children performed more actions than younger children. Through transforming the variable on a log scale, I could interpret results based on the magnitude between ages. Since the data I collected comprised of a large range of ages from 4 to 11 years old, the log transformation enabled me to see whether there were any changes with increasing age. However, this did not enable me to see at what specific age the change, if at all, happened. Time spent building was considered because not all children finished building at the same time and it was log transformed to account for the skewness of the data. Log transformation changed the slightly skewed data to a normal distribution. The participant's sex and what task they started with were coded as dummy variables (zero for females and one for males; zero for building task and one for divergent thinking tasks). The condition was coded as an index variable (one for instructions, two for no instructions and three for scaffolding).

All divergent thinking measurements were first centred by subtracting the mean of a variable from each value and standardized through dividing a centred variable by its standard deviation which hastened the computation and enabled easier interpretation of results (McElreath, 2016).

All response variables (all direct manipulations, exploratory behaviours, same and different manipulations) were counts and included many zero values. Individuals performed the task three times. Therefore, I analysed the behaviour data with multilevel zero-inflated models (ZIP). These are mixed models that use two probability distributions to model a mixture of causes. With the first probability distribution the model estimates the probability of a zero outcome. With the second probability distribution the model assumes Poisson distribution and its estimated mean (McElreath, 2016).

I explored two different predictors in the analysis, these were condition and divergent thinking measurements. For each of the predictor variable, I analysed four models. First, I built a "null" model in which I considered the age and sex of participant, I offset for their time spent building with blocks, and included varying effects, the individual ID and the build s/he was building. The second model included the condition and the third model the divergent thinking measurements. The fourth model included both condition and divergent thinking measurements. I used weakly informative priors.

In an additional analysis to set up the "null" model, I used a dummy variable that represented what participants experienced first, either the divergent thinking or building blocks task (variable code: F) and as a varying effect, the number of previous visits to the LSC (variable code: Mv) to see if these variables influenced the response variable. There were no differences between the models (Table 9). The differences in WAIC values were small and no model had a weight higher than 0.9. Therefore, I excluded these variables from the other analyses of behaviour measurements.

Table 9: Comparison of the models including parameter what did child begin with and how many times s/he visited museum in the past based on WAIC values and the Akaike weight.

	WAIC	pWAIC	dWAIC	weight	SE	dSE
mTIASB	2348.62	167.93	0.00	0.42	67.88	N/A
mTIASBF	2349.34	166.87	0.71	0.29	67.77	1.46
mTIASBMv	2349.37	169.06	0.75	0.29	68.03	2.12

3.2.5.2 Originality of the structures built by an individual and their divergent thinking score

Since the literature offers different views on divergent thinking as a predictor of success in innovation tasks (Beck et al. 2016; Runco and Acar 2012). I used divergent thinking measurements to predict the sum of the modes of the originality scores derived from ratings of the three structures individual built in the building block task, The response variable in linear models was the sum of modes of ratings assigned to the pairs of structures each individual built. Higher ratings (up to 21) would show very diverse structures whereas lower ratings (minimal was 3) very similar. The predictor variables I included were age and different divergent thinking measurements (Table 10). I did not include conditions as predictors, since there were no differences in ratings of the structures built by the same individual among the three conditions (results Section 3.3.1.4). I used weakly informative priors.

Table 10: Parameters include in the models predicting originality of the structures based on divergent thinking scores

model	parameters included in the models

m1.1	intercept + age
m1.2	intercept + originality
m1.3	intercept + originality + age
m1.4	intercept + fluency + flexibility + elaboration + originality
m1.5	intercept + fluency + flexibility + elaboration + originality + age

3.2.5.3 Ratings of the originality of the structures within individual

I used an ordered regression model to analyse the data from the ratings of the pairs of structures built by the same individual. The predictors included in the model were the participant's age, sex, time s/he spent building and the condition s/he was in. Since more than one rater evaluated the same pair of builds, the model intercepts also included a parameter defining the pair of structures. Intercepts were also the thresholds of the response variable which represent intercepts in ordinal categorical regression (Section 2.4.1). Each child built three builds which were used in the analysis and an index parameter defining the child was also included in the model. I used weakly informative priors.

3.2.5.4 Ratings within condition and between condition

When collecting the ratings for the pairs, both within conditions and between conditions, I showed participants only one example of the pairs of structures before collecting their responses. In order to make sure the raters did not need some time to establish their rating criterion, I checked the differences in their ratings given in the samples by excluding either the first 10, 20 or 30 answers and compared the basic model with the intercept and the rating responses. Since there were no differences in their posterior distributions, I used all of the answers from raters (Appendix 6).

I used an ordered regression model to analyse these data. The predictors included in the model were the participant's ages, sex, time spent building and the condition they were in. Since more than one rater evaluated the same pair of builds the model intercept also included a parameter defining that pair. Each of the children's structures were represented in pairs more than once, therefore I included in the analysis an index parameter defining the child as an intercept. I used weakly informative priors.

3.2.5.5 'Correspondence' between appearance and intended structure

I analysed the data of an adult's interpretation whether the child's structure resembled what the child has stated it was with an ordered categorical regression model. Each structure was rated from 6 to 18 times. In the analysis I decided not to include structures that children described as nothing and when they didn't assign any name to the structure (child answer: "don't know", "random"). The response variable in the model was the rating that the adults assigned to the structure. The explanatory variables were the children's age, sex, the condition the child was in, the sequential build (first, second or third) and the child participant index. I used weakly informative priors.

3.3 Results

3.3.1 Behavioural measurement

3.3.1.1 Direct block manipulations

I analysed how the conditions and the divergent thinking scores correlated with the number of manipulations participants performed during the building block task. The model including condition (mTIASBC, Table 11) had the highest Akaike weight, which indicates high probability of this model making the best predictions on new data in comparison to other models included in the analysis. However, the differences in WAICs are not large and together with the standard error of the differences, there is still a lot of uncertainty in deciding which is the best model.

Table 11: Comparison of the models predicting the number of manipulations of blocks including condition and divergent thinking scores based on WAIC values and the Akaike weight.

	WAIC	pWAIC	dWAIC	weight	SE	dSE
mTIASBC	2341.91	168.46	0.00	0.97	69.07	N/A
mTIASB	2348.62	167.93	6.71	0.03	67.88	4.83
mTIASBCDT	2360.29	186.86	18.38	0.00	71.25	11.64
mTIASBDT	2372.74	186.99	30.83	0.00	71.48	13.39

I explored the effect of the condition (instruction, no instruction and scaffolding) on the participant actions while building with blocks. To do this I built posterior estimates with the model including condition (Table 12). In all conditions the estimates crossed zero, therefore none of the conditions strongly correlated with the counts of block manipulations children performed while building (Figure 14).

Table 12: Estimates from the model including time, individual ID, age, sex, structure ID and condition.

	mean	standard deviation	lower 0.89	upper 0.89
βage_p	0.98	0.54	0.11	1.83
βage_l	1.49	9.28	1.04	1.94
$\beta male_p$	-0.24	0.57	-1.10	0.69
$\beta male_l$	0.12	0.14	-0.09	0.35
$\alpha instruction_p$	-0.39	0.58	-1.37	0.40
$\alpha instruction_l$	0.27	0.49	-0.46	0.99
αno instruction$_p$	0.62	0.61	-0.28	1.54
αno instruction$_l$	-0.16	0.49	-0.93	0.55
$\alpha scaffolding_p$	-0.10	0.54	-0.92	0.78
$\alpha scaffolding_l$	-0.15	0.49	-0.91	0.55

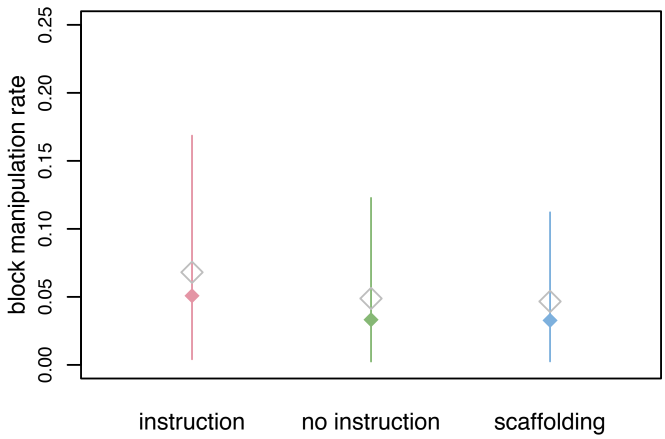


Figure 14: Effect of condition on rate of blocks manipulation by individual. Filled points present posterior mean, bars indicate an 89% credible interval and open points present actual mean from the data.

The probability of not performing any manipulations was slightly higher in the no instruction condition than in the direct instruction condition (Figure 15). Whereas, scaffolding (open questions) did not have any effect on the probability of not manipulating blocks. The average rate of manipulations was slightly higher in the instruction condition than in both the no instruction and scaffolding conditions (Figure 16).

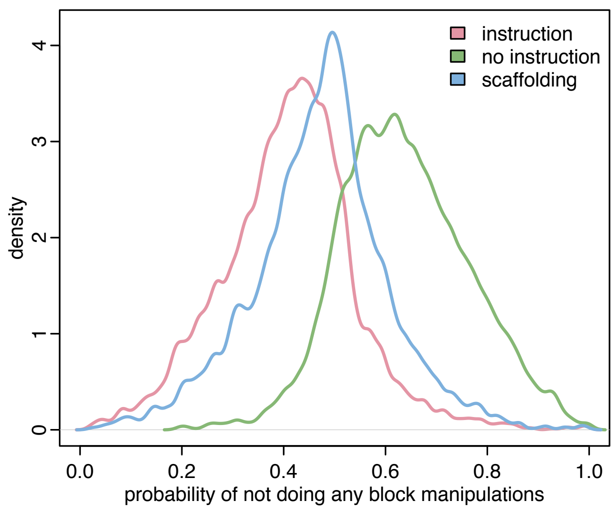


Figure 15: Posterior density plot of the probability of not doing any manipulations of blocks in different conditions.

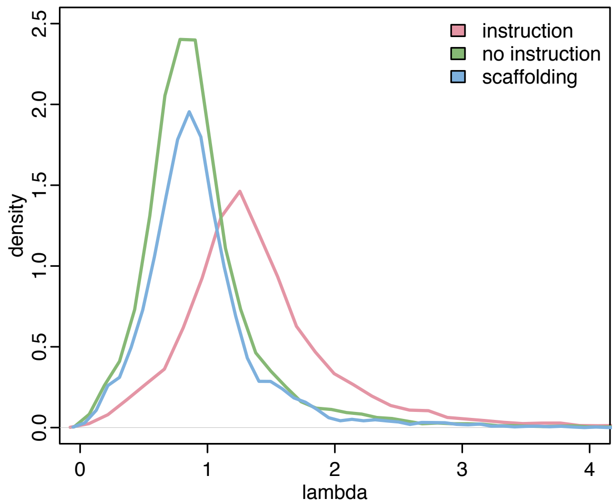


Figure 16: Posterior density plot of the average rate of manipulations in different conditions where age, sex, time building, individual and build are taken into account. Lambda is the average number of manipulations performed by participants.

The effect of sex of the children building with blocks did not differ between the conditions (for the estimates see Table 12; Figure 17).

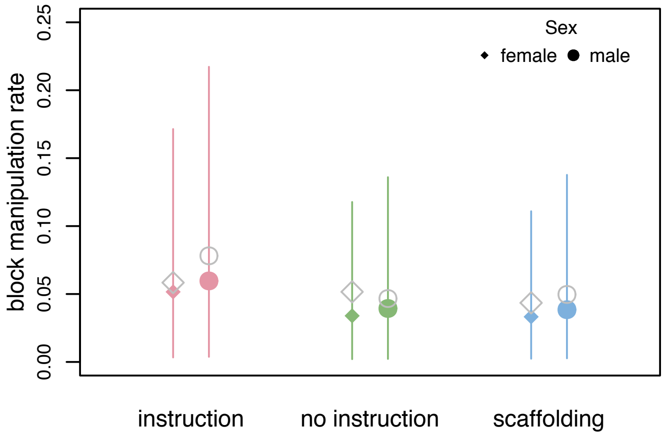


Figure 17: Effect of sex and condition on rate of blocks manipulation by individual. Filled points present posterior mean, bars indicate an 89% credible interval and open points present actual mean from the data.

Children's age was included in all models and did correlate with the number of block manipulations children performed. Older children performed more manipulations of blocks than younger (Figure 18).

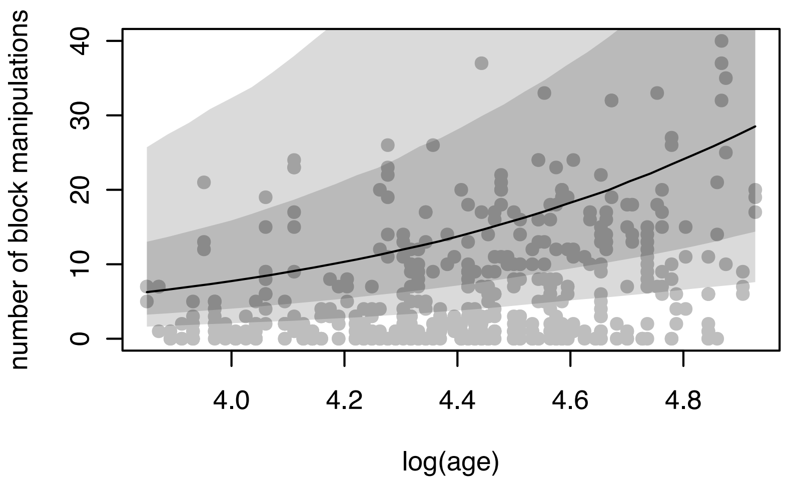


Figure 18: Posterior predictions of the effect of age effect on the number of block manipulations. Grey points are raw data, dark line is the median, the shaded regions from inside out: 67%, 89% percentile intervals.

The probability of not doing any manipulations of blocks was the same among the three structures (Figure 19). However, the average rate of block manipulations was greatest when children built their first structure (Figure 20).

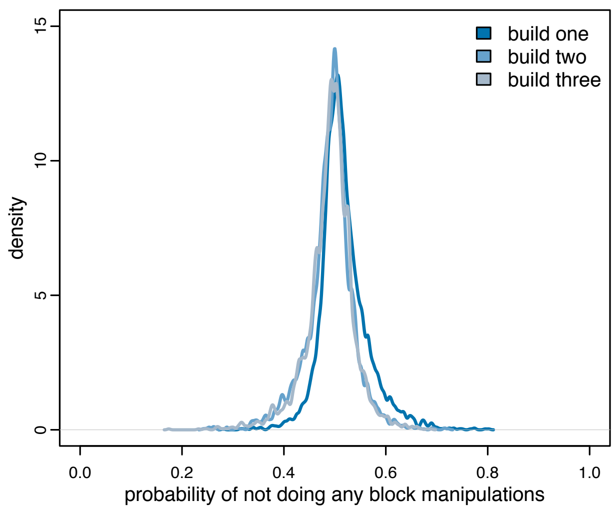


Figure 19: Posterior density plot of the probability of not doing any manipulations of blocks when building different builds (structures).

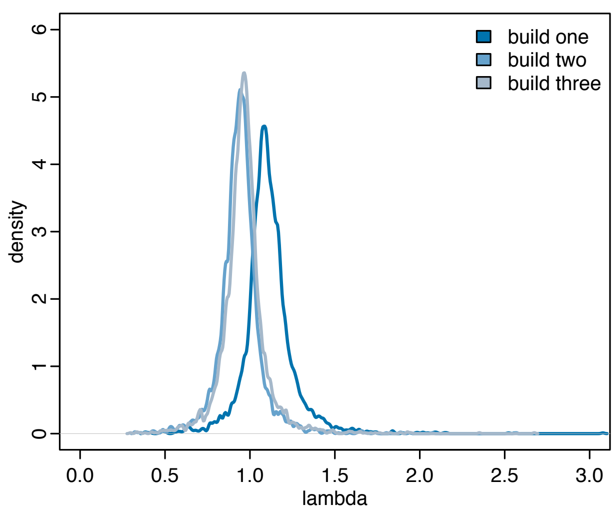


Figure 20: Posterior density plot of the average rate of manipulations when building different builds (structures).

3.3.1.2 Exploration

There was no effect of condition on whether participants explored the different ways they could manipulate the blocks and fit them into the structure they were building (Figure 21 and 22).

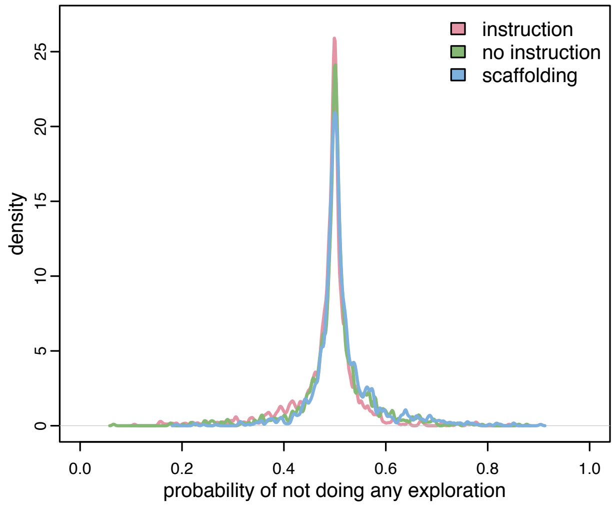


Figure 21: Posterior density plot of the probability of not doing any explorations when building in different conditions.

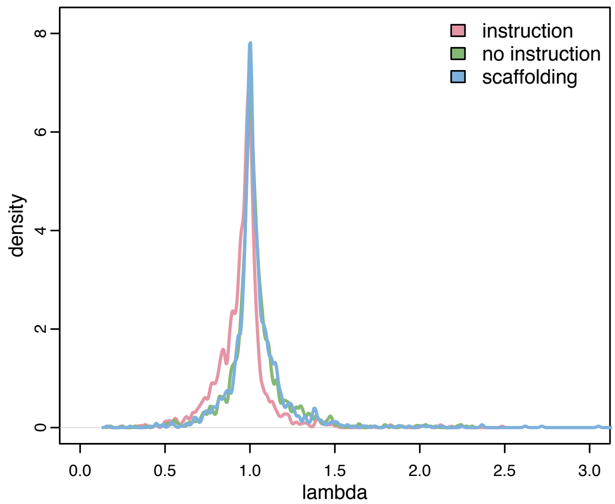


Figure 22: Posterior density plot of the average rate of explorations when building in different conditions.

3.3.1.3 Originality of the structures built by an individual and individual's divergent thinking score

The model including only the intercept and the age of individual best predicted the ratings of the structures individuals built (Table 10 and 13). Whereas, no measurement of divergent thinking correlated with the ratings the structures built by the same individual received. All posterior means estimates with 89% highest posterior density intervals from different divergent thinking measurements cross zero (Figure 23).

Table 13: Comparison of the models predicting the originality of the structures individual built including condition and divergent thinking scores based on WAIC values and the Akaike weight. See table 8 for description of the models.

	WAIC	pWAIC	dWAIC	weight	SE	dSE
m1.1	942.7	2.8	0.0	0.91	16.26	N/A
m1.2	947.4	5.8	4.7	0.09	16.80	1.76
m1.3	953.4	5.3	10.7	0.00	16.06	5.85
m1.4	953.9	13.5	11.2	0.00	17.14	5.60
m1.5	955.9	13.0	13.2	0.00	16.65	6.91

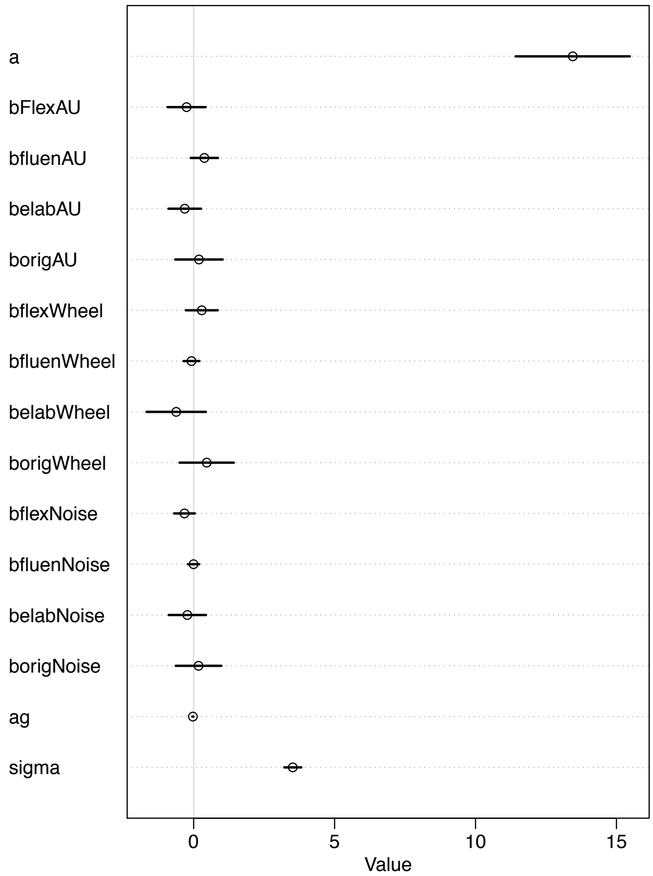


Figure 23: Posterior means of estimates and 89% highest posterior density intervals for the model including intercept, all divergent thinking measurements and age (Table 8, m1.5).

3.3.1.4 Ratings of the structure's originality within individual

Each participant built three structures in the same condition. These structures were compared with each other. I found no differences among the pairs of structures built by the same individuals across conditions (Figure 24). The most common rating of the pairs of structures the same individual built was 4 (1 – very similar to 7 – very different).

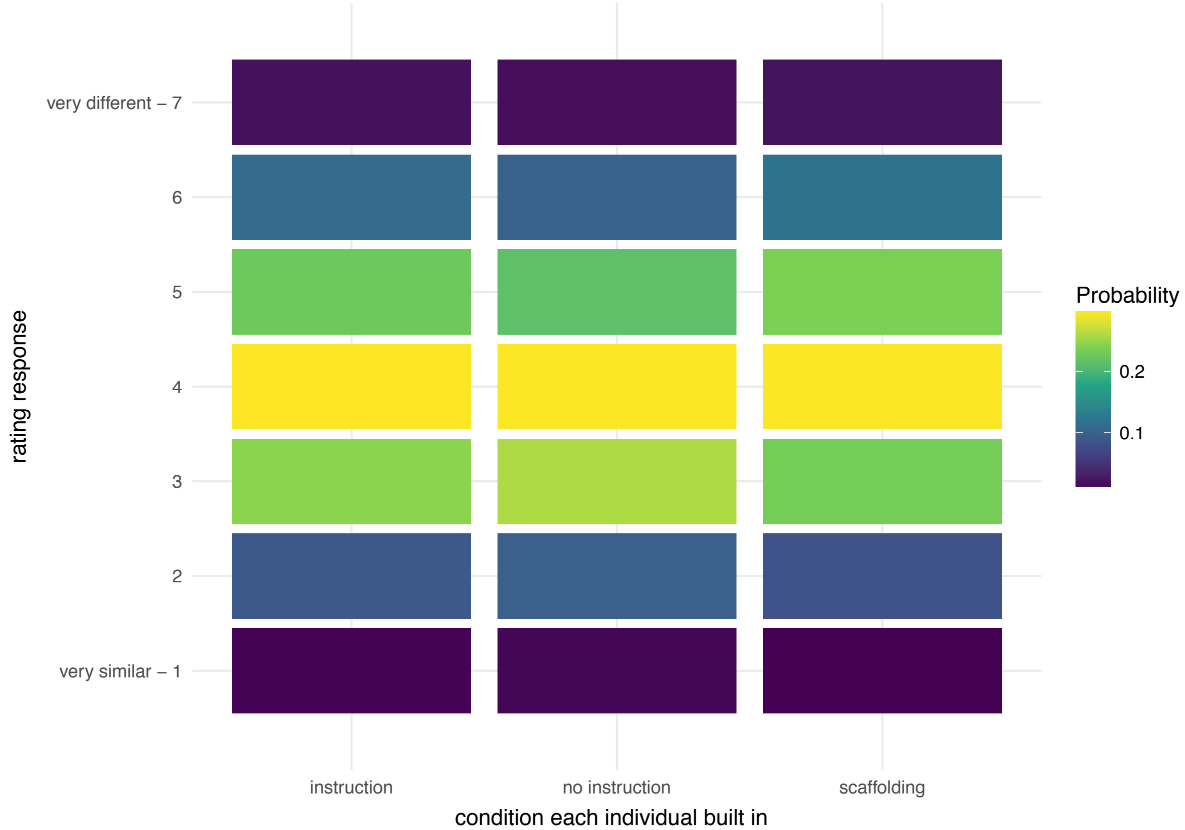


Figure 24: Marginal effects plot of the probability of the rating response (proxy for originality) of a pair of structures built by the same individual being rated on scale 1 – very similar to 7- very different.

3.3.1.5 Ratings of structure's originality within condition and between conditions

Each participant's first structure was compared either to the first structure from another participant in the same condition or a different condition. I was interested in the relation between conditions and the structure's originality. The condition did not have a strong effect on the originality of the structures (Figure 25). Pairs of structures in any type of comparison, within condition or between conditions, were most likely to be rated as either 6 or 7 (7 - very different).

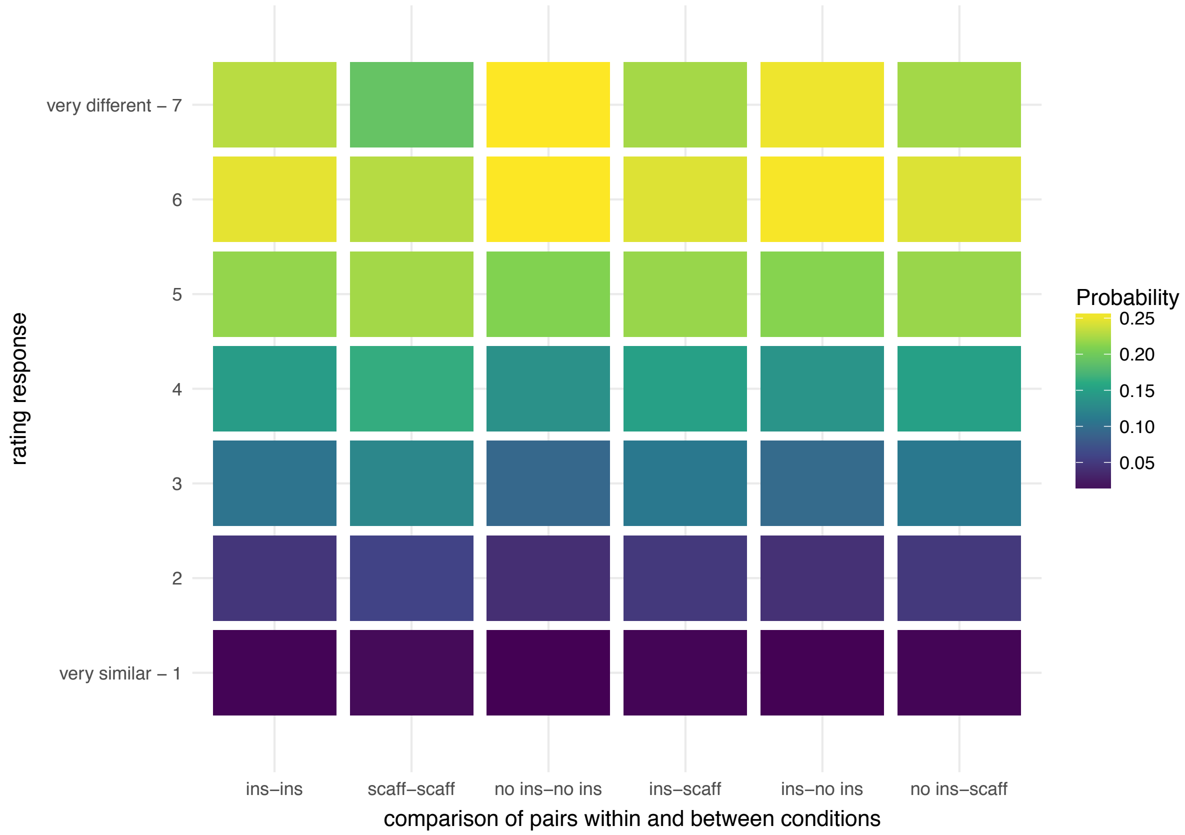


Figure 25: Marginal effects plot of the probability of the rating response (proxy for originality) of a pair of structures built by different individuals in either the same condition or in different conditions being rated on scale 1 – very similar to 7 – very different.

3.3.1.6 'Correspondence' ratings

The condition did not influence the correspondence of the structures to what children said they built (Figure 26). What children stated they built mildly corresponded with what the adult rater interpreted it to be (Figure 26). The highest probability of rating was 5 (1 – not at all to 7 – yes definitely).

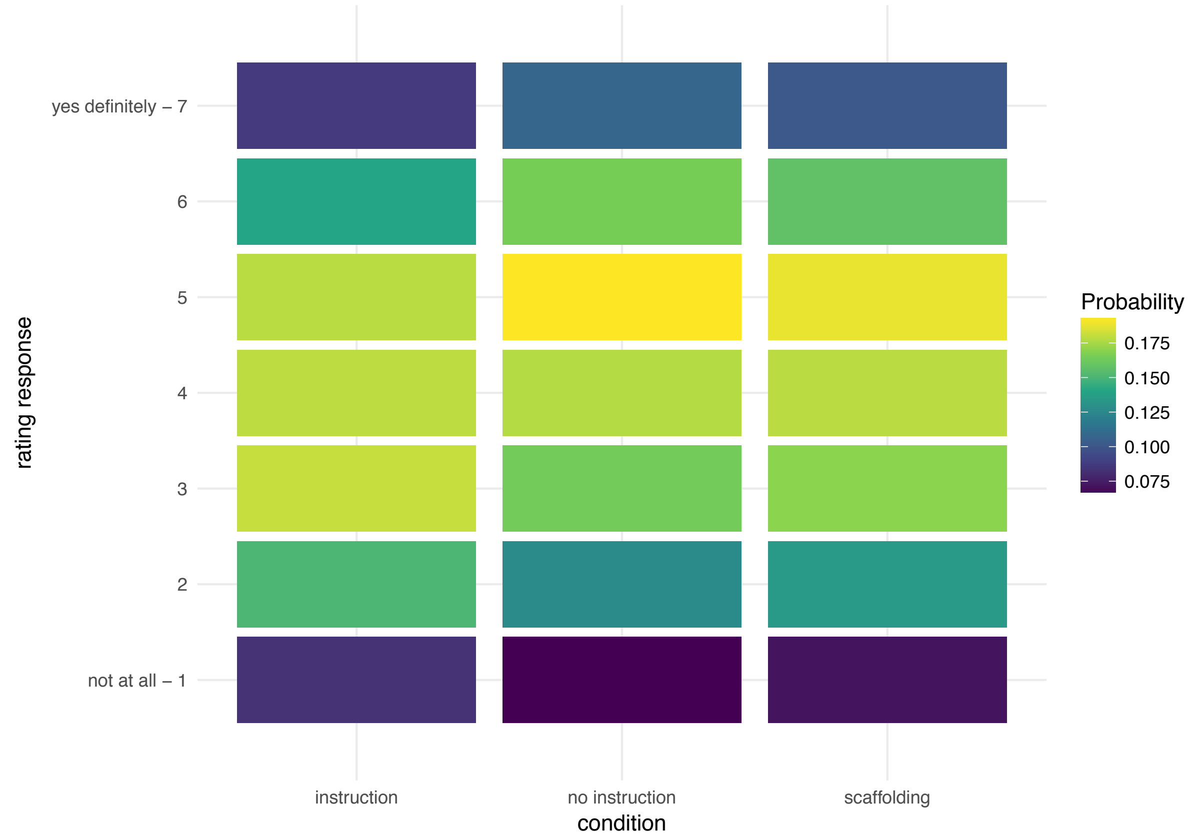


Figure 26: Marginal effects plot of the probability of the structure corresponding with what the child said she build being rated on scale 1- not at all to 7- yes, definitely.

3.4 Discussion

In this study I wanted to find out whether instructions (whether direct or guided) fostered or limited creativity in children. Guided instructions that are given as a demonstration of how to use something, and direct instruction (or teaching), are both forms of cultural learning. Furthermore, since the impact of instructions is challenging to test empirically, these kinds of questions have not been investigated so far. With the use of an open-ended task I wanted to empirically and quantitatively measure the impact of instructions on children's behaviours, and innovation, when interacting with an exhibit.

The main findings were that the building behaviours children expressed when interacting with the exhibit did not differ between the different conditions of direct instructions, guided and no instructions. Furthermore, the conditions that the children were in did not correlate with the originality of the builds children produced.

3.4.1 Building block behaviour across conditions

Contrary to my predictions, there were no differences in behaviours (counts of the manipulations of the blocks) when exposed to direct instructions, scaffolding and no instructions. The only small difference was in the number of manipulations when building with the first pile of blocks, which might result from priming (Kesek et al. 2011). In the instruction condition, children manipulated the blocks more when building for the first time than subsequent builds but this pattern was not observed in the other conditions. Children might have been more influenced by instructions at the start of the task because of their unfamiliarity with the blocks, in which case they would not explore them much but would remain loyal to already presented actions. The reason for why the effect of instructions on the building behaviour of children faded away after the first builds might also be that instructions were not strict enough. I did not ask them to use one of the demonstrated behaviours when building with blocks or tell them that this is the way it 'should' be done. The instructions were also not directed at building and were more of a demonstration of the blocks' affordances. I only demonstrated them what they could do with the blocks but did not imply that there was a "correct" way to use them. It could be that in more complicated tasks that would demand more skills, instructions would be beneficial (Clegg and Legare 2016a; Schillaci and Kelemen 2014). This is reflected in the literature where in some studies, such as goal oriented tasks, instructions and successful models were beneficial in aiding how children interacted with the task (Wood, Kendal, and Flynn 2015; Carr, Kendal, and Flynn 2015; Wilks, Collier-Baker, and Nielsen 2015) or the exhibit (Van Schijndel and Raijmakers 2016; Van Schijndel, Franse, and Raijmakers 2010; Haden et al. 2015). The latter was shown in the aforementioned study where children were exposed to successful analogical model builds (strong towers) of which they copied (Gentner et al. 2016). The diverse tasks used in research with children show that the picture about the benefits of instructing children is complicated and often depends on the type of tasks children are presented with (Dean and Kuhn 2007; Klahr and Nigam 2004; Bonawitz et al. 2011) and the context children are in (Rakoczy et al. 2009).

Building block behaviour was related to the age of a child building. Older children on average performed more block manipulations than younger ones. This might be due to the fact that blocks were heavy and not easy to manipulate (see section 3.6).

3.4.2 Explorative behaviours across conditions

Contrary to my predictions, exploration of blocks was neither limited nor facilitated in children exposed to different types of instructions or no instructions. This might be because blocks present an open-ended task but do not have many affordances which are present in the puzzle box tasks used in the past to study divergent thinking and exploration in children (Bijvoet-van den Berg and Hoicka 2014; Bonawitz et al. 2011). Where there are many features to the puzzle boxes and some even have additional novel objects which can be used to interact with the box (Bijvoet-van den Berg and Hoicka 2014), the wooden blocks used in this study only had one affordance. This was changing shape with the help of the rope connecting the cubes. In more open-ended and explorative tasks instructions hindered children's exploration (Bonawitz et al. 2011), the results in this study do not disapprove this but only add to it. The context and type of instructions and the task, namely exhibit, need to be taken into consideration when using exhibits tasks in a science centre (Van Schijndel, Franse, and Raijmakers 2010; Van Schijndel and Raijmakers 2016; Gentner et al. 2016; Haden et al. 2015; Corriveau et al. 2015). The aim of building blocks is to foster curiosity and creativity (Andy Lloyd pers. comm.) and not necessarily exploration. To understand how these three traits interact in children's learning and when exploring exhibits we would need to prepare more controlled experiments using diverse exhibits that aim to foster the development of different skills (Van Schijndel, Jansen, and Raijmakers 2018).

3.4.3 Originality of the structures built by an individual and her/his divergent thinking scores

In line with my prediction, the divergent thinking scores did not correlate with the ratings of the originality of structures individual built. This might be due to the fact that divergent thinking is not synonymous with creativity (Runco 1992; Runco and Acar 2012).

3.4.4 Originality of the final structures within individuals in each of the conditions

There were no notable differences among the structures built by the same individual in one of the three conditions. Mostly, the structures made by the same individual were rated as somewhere between very different to very similar regardless of the condition the participant was in. Children's divergent thinking (measured as the originality of the pairs of structures individuals built) was not facilitated or limited by direct instructions, scaffolding nor no instructions. This might be explained by the fact that also the ratings of originality of the structures individual built did not correlate with divergent thinking measurements. Children might also think they were restricted and did not want to explore different options of what to build. In comparison to the divergent thinking tests when children are asked to come up with different solutions and answers (Guilford et al. 1978; Wallach and Kogan 1965; Runco 1992), I did not prompt them to build diverse structures.

3.4.5 Originality of the final structures between individuals in the same condition

The first structures built by each individual were compared within the same condition and there were no differences between the conditions. Overall, the pairs of structures children built were likely to be rated as different independent of the condition they were built in. In a study, when children were exposed to affordances of the tools they were asked to innovate, they were more successful innovators (Neldner, Mushin, and Nielsen 2017). Therefore, I predicted instructions and scaffolding might have similarly enhanced the originality of structures children built. In contrast to this prediction, showing children specific affordances of the blocks (direct instructions) or fostering their discoveries of affordances (scaffolding) did not support or limit children's creativity when it came to putting blocks together into a final structure. Children's exploration can, in some contexts, be limited by instructions (Bonawitz et al. 2011; Wood, Kendal, and Flynn 2013a). The lack of originality of structures within direct instruction and scaffolding conditions might also be due to the fact children thought that there were 'correct' or conventional ways of putting the blocks together. This has been recorded in studies of children's functional fixedness, where children do not find a solution to a problem, because they think there is a conventional way of solving the task or using the materials available to them (German and Defeyter 2000; Alison Gopnik et al. 2017). Possible explanation might be attributed to the guidance being too vague. Furthermore, the originality of children's structures might be impacted due to the fact the model was an adult and children might follow age-model bias (Wood, Kendal, and Flynn 2012). It is already known that the context influences children innovation (Sheridan et al. 2016) and social learning (Burdett et al. 2016; Clay, Over, and Tennie 2018; Hanna and Meltzoff 1993; Wood, Kendal, and Flynn 2012). Innovation might have been suppressed when children were presented with social clues and models in their environment as in studies where children are shown solution to the problems and do not improve on them (Pinkham and Jaswal 2011; Carr, Kendal, and Flynn 2015; Reindl and Tennie 2018; Tennie et al. 2014). On the other hand, innovation might be fostered when children do not receive any instructions and are not exposed to social models.

3.4.6 Originality of the structures between individuals in different conditions

The first builds of each individual were also compared between different conditions. The differences between ratings of pairs between different conditions were negligible. This might be due to the fact the raters only provided ratings for the final structures and the direct instructions I provided children with might have more influence on the smaller parts of final structures. I only showed to ways of putting blocks together and did not model the final structure.

3.5 Limitations of the experimental set-up

The experiment had two constraints; very broad instructions and being challenging for younger children to manipulate of the blocks. First, the instructions children received were not strict and were only demonstrations of how to use blocks, but not limiting them in how to build final structures. The comparison of the final structures was a proxy measurement of creativity in children and it might be that there is no connection between the manipulation of the block, what condition children were in and the final structures built. Second, the blocks used in this study were large and not light. This might have impacted the behaviour measurements, especially in younger children who might not have performed as many manipulations as they would have done with a task that was easier for them to manipulate. However, all children were able to copy and repeat the experimenter's manipulations of the blocks when they were asked to change the block's shape, suggesting that the inability to manipulate the block would have a small impact with such a large sample size used in the study.

3.6 Conclusion

Direct instruction, scaffolding (open questions) and no instructions do not impact children's performance on the open-ended building block task. This might be due to several factors. First, poor choice of instructions that did not have a direct influence on the task from the beginning. I only showed children what can be done with blocks, but did not suggest what to build. However, when children were prompted to build high towers in a semi open-ended construction task, they also did not follow the models or build upon previous successful models they had been exposed to (Reindl and Tennie 2018). In addition, the similarity ratings of the high towers were low (Reindl and Tennie 2018), so the structures overall were not very similar to each other, which is in line with what I have found where in all conditions children did not build very similar structures to each other. Second, in an open-ended task where children were not limited by a goal, no type of instruction affected the building. This may have been cause I did not limit the space of solutions that could be built. For example, an instruction of building towers or castles, would probably have resulted in high similarity between structures compared to a control group in which children could build whatever they wanted.

To fully understand the impact of instructions on children's learning we need to combine studies of cultural learning, with educational research and studies in formal and experimental environments to studies in informal learning environments (Jipson and Sobel 2015; Callanan 2012). The literature on instructions impacting children's cognition and innovation already has conflicting results (Bonawitz et al. 2011; Dean and Kuhn 2007; Klahr and Nigam 2004) and science centres offer a great space to explore the question of using instructions and what type of instructions further creativity in children.

4 Is creativity impacted by asocial and social learning?

The influence of learning context on children's creative originality in a science centre

Abstract

Social learning is present in children from a very young age. Most studies explore it through dyadic experiments or through transmission chains and open diffusion. Here I use an exhibit within an informal learning environment to study social and asocial learning using an open-ended task. I predicted the structures children built in social learning conditions would be more similar to one another in comparison to the structures children built in asocial learning conditions.

I gathered data automatically with an Interactive Research Pod (IRP) which has been developed through the cooperation between academics and science centre practitioners. The IRP contained cameras and digital consent forms and asked children to build whatever they wanted from wooden blocks. I selected data from the video recordings taken from the IRP of 166 children aged between 4 and 12 years building in either social (children able to observe others building at the IRP through a transparent partition), asocial (children unable to observe others building at the IRP due to opaque partitions, but building at the similar time as others at the IRP) or asocial control (children building behind opaque partitions without anyone else being present at the exhibit) conditions.

What children said they built did not differ among conditions. Structures were less likely to be rated as very different to other structures (a measure of originality or creativity) in the social and asocial conditions than in the asocial control condition. Whereas in the asocial control condition, when no other models were present, the structures were more likely to be rated as original or creative.

Children are susceptible to social information and this study showed that social information influences creativity. Even so the differences among conditions were small. Nonetheless the results correspond to those of previous studies indicating that children are not very creative and are poor innovators, instead relying on social information. It is notable that this was the case in this study, even when the task had no goal and copying was not necessary.

4.1 Introduction

Several studies have shown children are susceptible to social information (Wood et al. 2016; Carr, Kendal, and Flynn 2015; McGuigan and Cubillo 2013; Flynn and Whiten 2010). Studies of social learning in children have investigated how they copy, whether they copy entire action sequences or just the end goals (Carpenter, Call, and Tomasello 2005; Carpenter, Akhtar, and Tomasello 1998; Meltzoff 1995; Whiten et al. 2009; Horner and Whiten 2005; Hopper 2010; Call, Carpenter, and Tomasello 2005; Tennie, Call, and Tomasello 2006), when and why they over-imitate (Lyons, Young, and Keil 2007; Lyons et al. 2011; Evans et al. 2018; Nielsen, Moore, and Mohamedally 2012; Keupp, Behne, and Rakoczy 2013; Keupp et al. 2015; Clay and Tennie 2017), and who they copy (for a review see Wood, Kendal, and Flynn 2013b). Mostly studies use dyadic experiments, where the demonstrator (usually an adult) performs a set of actions that have a goal which leads to a reward and then the child is asked to solve the same task. In cases when researchers are interested in whether children copy the majority of others or if the gender of the demonstrator influences their social learning they might also show children a video of the demonstration (Morgan, Laland, and Harris 2015). Besides studies of social learning in dyadic experiments, studies of transmission chains can also be found in the literature. To this point there have only been several studies of this type (Flynn and Whiten 2010) and only a few studies used open diffusion methods in children (Flynn and Whiten 2010; McGuigan et al. 2017; Dean et al. 2012).

The methods used in diffusion studies include several methods. These being; diffusion chains, where the information or a behaviour is transmitted down a chain of individuals, like in the game 'telephone' (Horner et al. 2006; Flynn and Whiten 2008; Flynn 2008; McGuigan and Graham 2010; Hopper et al. 2010; Tennie et al. 2014), replacement method, where naive individuals are introduced to the chain (Caldwell and Millen 2008a) and open diffusion where individuals can openly interact in the chain (Flynn and Whiten 2012; Whiten and Flynn 2010; McGuigan et al. 2017; Dean et al. 2012). Diffusion chain experiments are especially useful for observing the transmission of traditions and knowledge through groups of children (for a review in adults see Mesoudi and Whiten 2008) and exploring cultural ratchets (Tennie et al. 2014) of which supposedly lead to an accumulation of knowledge and supports cumulative cultural evolution. However, diffusion chains and replacement diffusion are still limited to pairs of interacting children, therefore a method of open diffusion (Flynn and Whiten, 2010) is used to explore how knowledge spreads more realistically through groups of children. The latter provides higher ecological validity at the expense of lower experimental validity and the data are messier therefore, a balance must be reached.

Studies using diffusion chains mostly explored how and when information was transmitted between children. For example, whether children copied the means or the goals (Horner et al., 2006), whether over-imitation persisted throughout the chains (Flynn, 2008; McGuigan and Graham, 2010), whether the sex of individuals in chains affected their transmission (Flynn and Whiten, 2008), the presence of social learning biases (Flynn and Whiten, 2010) and how long the demonstrated method for a solution to the task persisted in generations along the transmission chain (Hopper et al., 2010). Only recently a few studies investigated innovation in children when transmitting information (Reindl and Tennie 2018; Tennie et al. 2014; Dean et al. 2012; McGuigan et al. 2017). Through transmission chains they found that children could invent and transmit more efficient solutions to tasks when an inefficient solution was seeded in the beginning of the chain, suggesting innovation (Tennie et al., 2014). However, children did not improve on already efficient solutions, but copied them. In the case where children were presented with arguably open-ended task, in which they had to build high towers from spaghetti and plasticine, the similarity among towers children in same transmission chain built was greater in comparison to those among different chains (Reindl and Tennie 2018).

Building on the methodology of open diffusion, experiments using micro-societies in children were developed (Whiten and Flynn, 2010). These were primarily developed to explore the connection between social learning and innovation. In micro-societies a knowledgeable model aware of the solution to a task is seeded in a group of naive individuals and researchers observe how this solution spreads throughout the group. This enables researchers to observe transmission and, in some cases, also inventions of new traditions (Whiten and Flynn, 2010). These micro-societies permit the study of social learning together with innovations in groups and therefore offer an exploration of cumulative culture in children. In a comparative study of groups of children, capuchin monkeys and chimpanzees, children were able to reach high-level solutions to tasks through direct teaching, imitation and pro-sociality (Dean et al., 2012). For example, when children in groups were presented with a task which required new inventions to solve, diverse types of social learning together with novel solutions emerged in the group of children (McGuigan et al., 2017). Besides diffusion studies, recently, social network analysis has been implemented to study children in groups (Hanish and Rodkin 2007; Golemiec et al. 2016; Rawlings, Flynn, and Kendal 2017) and transmission within social networks.

The majority of studies exploring social learning, imitation and innovation in children were performed in nurseries, schools and laboratories (Carr, Kendal, and Flynn 2015;

Wood, Kendal, and Flynn 2012, 2013a). In some cases, researchers recruited children in science centres and museums, but the studies were still conducted in a controlled setting (Reindl and Tennie 2018; Reindl et al. 2016). Social learning is explored together with innovation in studies that investigate cumulative cultural evolution (Tennie et al. 2014; Reindl and Tennie 2018; Dean et al. 2012). A few of these studies used open-ended tasks where children were asked to build high towers out of spaghetti and plasticine (Reindl et al. 2017; Reindl and Tennie 2018). I would argue that these were not entirely open-ended since there was a clear goal; building high towers.

Instead of selecting a task with a clear goal and solution, I decided to explore creativity through an open-ended task in a science centre. Creativity enables innovation and inventions together with other processes, such as explorations and neophilia (Carr, Kendal, and Flynn 2016). In the literature of science centres, it is known that visitors spend a longer time engaging with open-ended exhibits than exhibits that are limiting (for example where visitors only read about the exhibit or lift a flap to find answers) (Sandifer, 2003). The longer people interact with open-ended exhibits the more likely they are exposed to other's methods of interacting with the task such as their building processes, thus creating the perfect conditions to study social learning and innovation.

In this study I explored children's behaviour in an informal learning context which could be seen as a natural environment for children. I conducted the study without the researcher being present which removed the possibility of affecting how children would behave if an 'authority' was present (Whiten et al. 2016). Their main task was to enjoy, explore and play with exhibits. To test how social learning is expressed in a context that is not goal oriented, open-ended tasks are well suited. In this study I explored whether unnecessary imitation also occurred in a completely open-ended task and whether some traditions and novel structure styles emerged in any of the conditions, social and asocial. In this study I used a building blocks task in which children were asked to build whatever they wanted. A building blocks task is an open-ended task with 100 of the same shaped wooden blocks. The task was selected as it does not limit children in their creativity and does not pose a problem to solve which is more common in other studies of innovation in children (Beck et al. 2011; Nielsen 2013), and by the age of four children have already developed some spatial cognition skills (Reifel 1984).

I explored whether the originality of the structures children built were influenced by the opportunity to see another child building or another's final structure. To do that I collected data through the use of an Interactive Research Pod in three conditions: social (children able to observe others building at the IRP through a transparent partition), asocial (children unable to observe others building at the IRP due to opaque partitions, but were building at the similar time as others at the IRP) or asocial control (children building behind opaque partitions without anyone else being present at the exhibit). I predicted social conditions would limit children's creativity (Reindl and Tennie 2018), the originality score would be lower than in asocial condition. Whereas, asocial conditions would promote it. I predicted children would imitate others building in the social condition, especially when building at the same time, since children mainly imitate actions of the demonstrator and not just results from demonstrators actions (Whiten et al. 2009; Horner and Whiten 2005; Hopper 2010; Call, Carpenter, and Tomasello 2005; Tennie, Call, and Tomasello 2006).

4.2 Methods

The methods are divided into four sections. First, I introduce the participants, apparatus (the Interactive Research Pod), the procedure of selecting data from and a novel method to rate pairs of builds based on their similarity. Second, I present a statistical analysis and the results of the inter-coder reliability, burn-in period of the ratings and the position of the images that appear on the screen in the rating app. Third, I present the statistical analysis with which I explored the data from individuals building at the Interactive Research Pod. Fourth, I present the statistical analysis of the data from the originality of pairs of structures that also include behavioural measurements and the combined data from pairs of individuals.

4.2.1 Participants

The study was conducted in the Brain Zone of a science centre, Life Science Centre in Newcastle, UK. The data were collected between the 15th of October 2016 and the 23rd of February 2017. I collected data from 166 participants (84 males, from 4 to 12 years old). I gathered 60 (26 males) individuals in a social condition and 106 individuals (48 males) in an asocial condition (66 in asocial and 40 in control asocial) (Table 15). From these I formed 68 pairs in the social and 56 pairs in the asocial conditions, 20 of which were pairs in the asocial control condition (Table 16). The pairs were formed from children that visited with their guardians.

Table 14: Explanation of the three conditions

Social condition	Partitions	Transparent
	Pairs	Children built at the same time or the same final structure or model was present at the research pod
Asocial condition	Partitions	Opaque
	Pairs	Children built at the same time or the same final structure or model was present at the research pod
Asocial control condition	Partitions	Opaque
	Pairs	Children built alone when nobody else was building at the Interactive research pod on different days

Table 15: Number of individuals in each condition

Condition	Number of individuals	Females	Males	Age range (in years)
Social	60	34	26	4 – 12
Asocial	66	37	29	4 – 12
Asocial control	40	21	19	5 – 12

Table 16: Number of pairs in each sex category

Pairs condition	Number of pairs	Female pairs	Male pairs	Mixed sex pairs
Social	68	20	16	32
Asocial	36	11	6	19
Asocial control	20	5	6	9

Table 17: The profile of individual participants

Condition/Age	Social	Asocial	Asocial control
4	1	3	0
5	7	3	5
6	7	10	5
7	5	11	10
8	6	8	4
9	9	7	6
10	10	7	6
11	10	7	1
12	5	10	3

4.2.2 Apparatus

The Brain Zone has numerous exhibits which introduce visitors to behavioural experiments. One of them is the Interactive Research Pod exhibit which was developed through the collaboration between researchers from Durham University and practitioners at the Life Science Centre (Kendal et al., 2016, Rudman et al., 2017).

The Interactive Research Pod offers a creative activity where children are asked to make a structure out of wooden blocks. The exhibit comprises of a round table with three separate stations to build and adjustable partitions that divide the three stations which can either be opaque or transparent. Cameras and table computers are mounted onto the partitions to record the participant’s behaviour, gather general information about them and gain their assent to participate in the study. Each station has approximately 100 pieces of wooden blocks (Keva planks). These are brown coloured cuboid wooden blocks of the same size. Each block was 6.35 mm thick, 19.05 mm wide and 114.3 mm long.

4.2.3 Procedure

The experiment was conducted without any researchers being present at the exhibit. The use of digital tools embedded in the exhibit enabled me to collect all data automatically.

4.2.3.1 Children building

Each child approached the Interactive Research Pod and they or their guardians completed the consent form and questionnaire at the station before the child started building. After that, the child could build for as long as they wanted. When they finished, children had to answer the question “What did you build?” and guardians had to agree again that their children’s data could be used for research purposes.

4.2.3.2 Selection of the data from the Interactive Research Pod

To select the data used in this study I first evaluated the data from the consent forms which were collected through tablet computers that were attached to the Interactive research pod. I only accepted the data when parents gave consent to use them and when the participant was within the age range between 4 and 12 years old.

After an initial selection of consent form data, I identified each participant on the video. Since there was not a researcher present to control the environment, many of the children whose parents gave consent built with the help of a parent or in groups with other children. I did not use these data. I only selected children that were building alone at one station of the Interactive Research Pod. For both the social and asocial conditions, I selected pairs of children that were either building at the same time, shared the same model or were both exposed to a final structure present on the pod (for a detailed description see Table 18). In the social condition the partitions of the Interactive Research Pod exhibit were transparent and in the asocial condition the opaque partitions were used (Table 14).

Table 18: Ethogram of behaviours measured at the Interactive Research Pod

Same time	Yes – 1 No – 0	at least for some amount of time both participants in the pair were building together on different stations at the Interactive Research Pod
Same model	Yes – 1 No – 0	participants in the pair could both see the same person (child, adult or group) building at the station that none of the participants constituting the pair are not building on at least for some amount of the time while building
Same build	Yes – 1 No – 0	participants in the pair were exposed to at least one final structure at one of the building stations while building. For example, in the case when the first participant (A) built at station one and left their structure standing, the second participant (B) observed it before destroying it to start building a new one

Three children could build at the same time at the Interactive Pod. However, children did not usually start or finish at the same time. In these cases, there were chains of children sharing the same final structures and building at the same time directly or indirectly. For example, ideally children A, B and C started at the same time, finished at the same time and therefore, they formed pairs. In a situation when child A at station one started alone, child B joined at station two and started building. Then child A finished and afterwards child C started building on station three of the pod. These children could all be paired together (these pairs would be A-B, A-C, B-C.) because they either built at the same time or observed the same finished structure. When child A built at station one but did not give assent to use their data and two other children (child B and C) started building, these two children (B and C) were paired together as they could see the same child building (child A) even if they were not building at the same time. If children A and B built at the same station one after another and B had seen the final structure child A built, these two children were also paired together as sharing the same final structure.

Children in pairs were also part of a larger cluster of individuals building at the Interactive Research Pod in a similar time frame. In cases when all data from consent forms and videos were available for a larger cluster of children, I gathered data as a cluster in the social condition. These clusters do not necessarily present clear transmission chains, since there were actors present at the pod that I could not include in the analysis because they were either adults, children whose guardians did not consent to the study or groups of individuals. All together I gathered 8 such clusters in the social condition. The cluster sizes varied from 3 individuals up to 7.

I followed the same rules for combining pairs in the asocial condition. The pairs in the asocial condition were selected when partitions of the Interactive Research Pod exhibit were opaque. However, I added a control for the asocial condition because children could observe others building when they approached the exhibit. The control included 20 pairs from children that built alone at the table when the partitions were opaque but I paired them with children in the same condition but on a different day.

When I identified all the pairs to compare I extracted images of the children's final builds from the video recordings of the Interactive Research Pod. Adult participants rated the structures from the images based on their similarity.

4.2.4 Ratings of the pairs of structures

To evaluate the pairs of builds in each condition I followed methods presented studies of cumulative culture (Caldwell and Millen 2008b; Reindl and Tennie 2018). I assigned the images into pairs and I created a web application in the java script programming language using jsPsych library (de Leeuw 2015) to compare the pairs based on their structural similarities. I recruited 103 participants (52 males; median age: 28) through Prolific. Each participant rated 124 pairs. They were asked "How similar are the structures?" and had to rate them on a scale from 1 (very similar) to 7 (very different). The pairs were presented in random order for each participant to rate. Each pair was sometimes shown in an initial order (pair A and B: left - image A and right - image B) and sometimes in reverse order (pair A and B: left - image B and right - image A). Each participant spent around 15 minutes to rate the pairs and received compensation of £1.50.

4.2.5 Scoring the "What did you build?" measurement

When participants finished with building, they were asked what they had built through the digital consent form present at the exhibit. I rated this data based on the originality of the answers and the amount of details included in the children's' answers. I summarised the answers provided by the children into categories based on the most common description. For example, if they said "house with doors" I selected house as the main subject and doors as a detail. I then counted the responses and divided them into three categories: common, less than 5% and unique. Common were subjects that more than 5% of participants stated they built, less than 5% included all subject that at least 2 participants stated they built and unique subjects were the structures only one child said s/he built. Details were rated based on any amount of elaboration they added to the main subject. Details present a binary variable with a value of 1 if there was any elaboration and a value of 0 if there was no elaboration. For example, answers “garden for horses and cows” would score 1 for elaboration, while just “garden” would score 0.

4.2.6 Statistical analysis

All statistical analyses were carried out in R 3. 3. 2. (R Core Team 2013) using the *rethinking* package (McElreath 2016). Posterior estimates were generated using RStan 2.17.3 (Stan Development Team 2018). Inter-coder reliability with Cohen’s Kappa coefficient has been calculated using the *irr* package (Gamer, Lemon, and Singh 2019) in R.

4.2.6.1 Inter-coder reliability

To evaluate the reliability of the coders an additional 25% of the randomly chosen behavioural data were coded simultaneously by two researchers one being unaware of the conditions and the predictions. The variables of interest were whether the children built at the same time, had the same model and same build as well as whether the answers to “what did you build?” included elaboration or not. A Cohen’s kappa of .967 was achieved, demonstrating almost perfect reliability.

As an additional test to classical Cohen’s kappa following Bayesian statistics methods I then compared WAIC values and their weights between an intercept only model and a model including coder ID as a predictor. All models were binomial, had weakly informative priors and were approximated by a Gaussian distribution with the ‘map’ function in the *rethinking* package. There was no difference between the intercept only and coder ID models (Table 19). Therefore, the coding was reliable.

Table 19: Models of inter-coder reliability. M1.1 is the intercept only model, m1.2 is the intercept and coder ID model.

	WAIC	dWAIC	weight	SE	dSE
m1.1details	109.64	0	0.72	4.77	N/A
m1.2details	111.57	1.93	0.28	4.92	0.15
m1.1sameTime	102.87	0	0.74	6.94	N/A
m1.2sameTime	104.95	2.08	0.26	7.05	0.50
m1.1sameModel	114.59	0	0.72	2.27	N/A
m1.2sameModel	116.45	1.85	0.28	2.25	0.10
m1.1sameBuild	112.50	0	0.73	3.47	N/A
m1.2sameBuild	114.55	2.05	0.27	3.56	0.36

4.2.6.2 Initial and reverse position of images in the pair

Images in pairs were shown to adult raters in different orders and on different positions of the screen. I used a sample of 40 pairs to evaluate whether the position of the image in the pair influenced the ratings of the pair's similarity, I compared a null model with the intercepts as pair ID and a model with an intercept; pair ID and a predictor; position of the images in the pairs (initial = 0, reverse = 1). I used weakly informative priors and the ordered categorical models were fitted with MCMC using 'map2stan' function. Positions of the images in pairs did slightly influence the ratings (Table 20). When pairs were shown in reverse order to the initial one there was a lower probability of being rated on higher categories on the Likert scale (Figure 27). I therefore assigned randomly images to the left or right position of the screen.

Table 20: WAIC values of the models of the position of the pairs. M1.1position is the model with intercept with pair ID and m1.2 position is the model with intercept pair ID and the predictor position of the images in the pair.

	WAIC	dWAIC	weight	SE	dSE
m1.2position	12075.4	0	1	92.00	N/A
m1.1position	12097.3	21.9	0	91.43	10.13

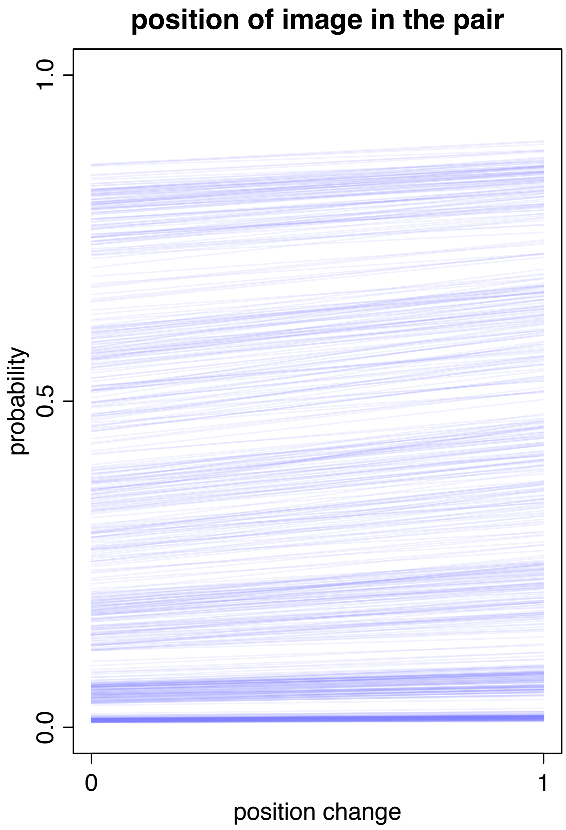


Figure 27: Posterior predictions of the ordered categorical model with the predictor the position of images in the pair. 0 presents the initial position (e. g. A-B), while 1 present the reversed positions (e. g. B-A). The lines and gaps between the lines present the ratings from 1 to 7. The plot presents how the distribution of predicted responses varies by the position of the images in the pair.

4.2.7 Individual data analysis

I analysed data from individuals in different conditions to see whether what they said they built correlated with any of the predictors, such as age, sex and condition.

4.2.7.1 What did you build?

I assigned three categories; common, less than 5% and unique to the descriptions of children's builds. I changed each of these categories into binary variables, with that I created three dummy variables to analyse the three-state response variables. I created three different response variables (common, less than 5% and unique) and ran three separate analyses for each of the responses. In the binomial models I included predictors age, sex and condition. I used weakly informative priors.

4.2.7.1.1 Wordclouds

I created wordclouds using the *wordcloud* package in R, to present the different frequencies of the answers to the question "what did you build?".

4.2.7.1.2 Details

Some answers the children provided to the question "What did you build?" included elaboration. I used binomial models with response variable; details (details in the answer = 1, no details = 0) and predictors; age, sex and condition. I used weakly informative priors.

4.2.7.2 Analysis of the ratings of pairs of structures

I analysed the ratings assigned to the pairs of builds by the adult raters with ordered categorical models. I gathered a large amount of ratings data altogether after excluding the pairs that were not rated by some individuals, there were 12452 rates in the final data set. I used the mode as a representation of the originality 'rating' each pair received by many independent raters. Since ratings are Likert items represented on a Likert scale these response categories have a rank order and since the intervals between them are not considered equal, they cannot be treated as continuous variables. Therefore the median or mode is used to present the central tendency of the data (Jamieson 2004). The mode presents the most commonly assigned value to the pair and the method of rating pairs of images was developed to evaluate the similarity of pairs objectively with the use of a large number of raters.

4.2.7.2.1 All conditions

First, I analysed whether there was a correlation between the response variables, in this case, ratings of the pairs with sex category (female-female, male-male, female-male), average age of the individuals in the pair and absolute difference in ages between individuals in the pair and the conditions (social, asocial and asocial control) they were building in. I averaged the age of the individuals in pairs to enable me to see if pairs with older or younger children received different originality ratings for their builds. I also computed the absolute difference in their ages to account for the magnitude of the age gap which is not represented by averaging the ages of the individuals assigned to pairs.

I ran four ordered categorical models. The first included the sex category, average age and absolute difference in age. The second included the sex category, average age, absolute difference in age and the conditions. The third included the sex category, average age, absolute difference in age, the interaction between average age and absolute difference in ages and the conditions. The fourth included only the intercepts and the conditions. I used weakly informative priors.

4.2.8 Social and asocial – same time, same build, same model

I subset the rating data to only include two conditions, social and asocial, to see whether the variables; building at the same time, being exposed to the same structure or sharing the same model could predict the originality rating assigned to the pairs of structures. I ran six models with weakly informative priors. The first included the sex category, average age, absolute difference in age and the interaction between average age and absolute difference in ages. The second included the sex category, average age, absolute difference in age, the interaction between average age and absolute difference in ages and the conditions. The third included the sex category, average age, absolute difference in age, the interaction between average age and absolute difference in ages, the conditions and the behaviours (Table 14 and 18). The fourth included the sex category, average age, absolute difference in age, the interaction between average age and absolute difference in ages and behaviours. The fifth included only behaviours. The sixth included conditions and behaviours.

4.2.8.1.1 Social – cluster effect and same time, same build, same model

I subset the originality ratings data to only select the data in the social condition and compared within cluster ratings of the pairs with pairs that were not in any of the clusters. Clusters were all joined in one binary variable (no cluster = 1, any cluster = 0). I ran three ordered categorical models. The first included predictors of building at the same time (yes = 1, no = 0) and being in a cluster. The second included only one predictor; building at the same time. The third included only one predictor; being in a cluster.

I ran an additional ordered categorical model where the predictors were building at the same time, being exposed to the same build present at the Interactive Research Pod and sharing the same model building at another station. This enabled me to explore whether pairs that were building at the same time, were exposed to the same final build or shared the same model were rated as more similar.

I ran an additional model to see whether pairs belonging to clusters had similar ratings. I used an ordered categorical model with the intercept of clusters. The pairs that did not belong to any of the clusters formed their own cluster.

4.2.9 Heatmaps

As an additional description of the cluster data in the social condition, I prepared heatmaps with *ggplot2* package (Wickham 2016) in R. I have created heatmaps that show ratings of the comparison between pairs of the structures that children built at a similar time. These heatmaps represent the pairs that comprised the same cluster in the social condition and the mode of ratings for a pair. I did this to see whether there were any trends emerging in what children built when in the social condition at the IRP. I included the table describing the clusters and the sequence of the events when participants in the clusters were engaging with the IRP.

4.3 Results

4.3.1 Individual data analysis

4.3.1.1 What did you build?

Children did not classify their structure any differently in any of the conditions. All models did not differ from the intercept-only model (Table 21, Figure 28). In the models predicting the common and less than 5% category, the uncertainty of the prediction was large.

Table 21: WAIC values of models with the categories from "what did you build" answers. M1.1 is intercept-only model, m1.2 is intercept, age, sex model and m1.3 is intercept, age, sex and conditions model.

	WAIC	dWAIC	weight	SE	dSE
m1.1common	229.2	0	0.81	3.51	N/A
m1.2common	233.3	4.1	0.11	3.58	0.61
m1.3common	233.7	4.5	0.09	5.48	4.22
m1.1less_than_5p	199.9	0	0.70	10.95	N/A
m1.2less_than_5p	202.2	2.3	0.22	11.51	2.98
m1.2less_than_5p	204.3	4.4	0.08	12.08	4.36
m1.1unique	187.6	0	0.82	12.55	N/A
m1.2unique	191.2	3.6	0.14	12.96	1.23
m1.3unique	193.7	6.1	0.04	13.44	3.16

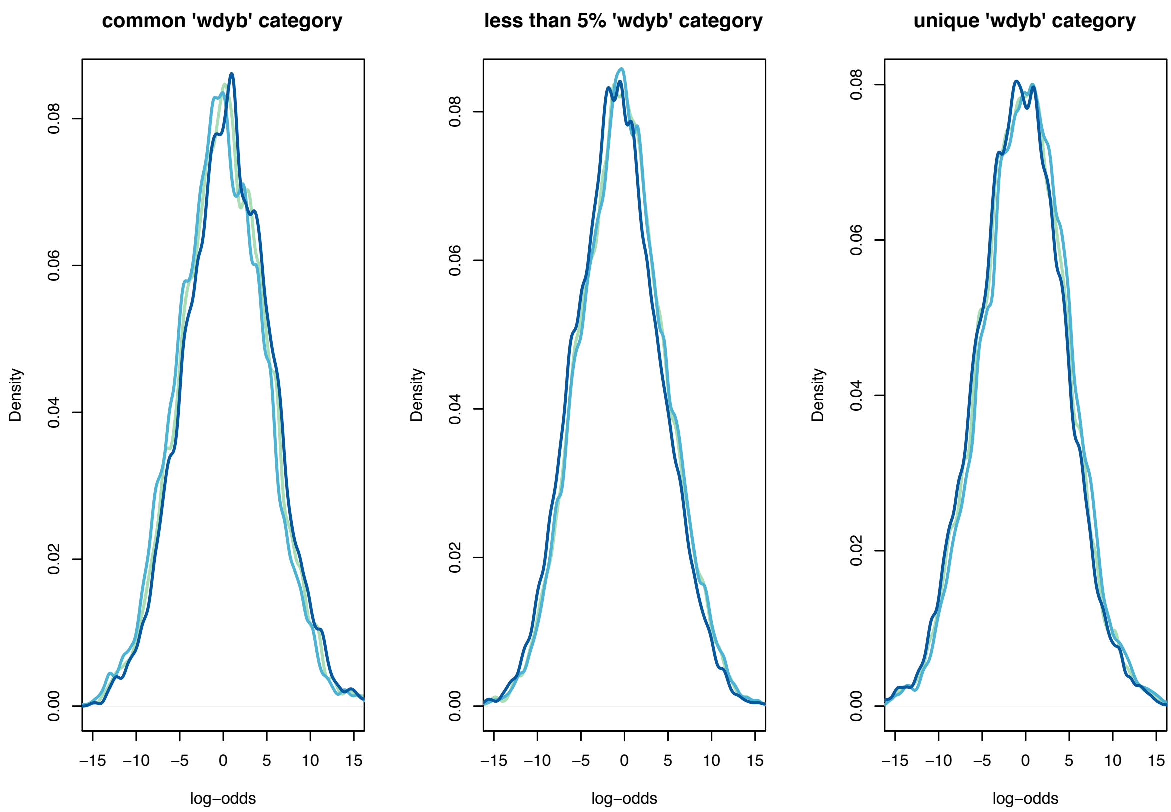


Figure 28: Log-odds of conditions from the models predicting the category of the answers to the "What did you build?" question (m1.3 including age, sex and condition). Dark blue – social condition, green – asocial and light blue – asocial control.

4.3.1.2 Wordclouds

The most frequent builds were house and tower in all conditions (Table 22; Figures 29, 30, 31).

Table 22: Counts of the answers to the question what children said they built

What did you build?	Count of "What did you build?"
aquarium	1
bank	1
base	1
basket	1
bed	1
box	1
bridge	1

building	2
bungalo	1
bunker	2
castle	7
cave	1
chairs	1
church	1
cinema	1
colloseum	1
cottage	3
crisscross	1
engine	1
fence	1
fort	3
gallery	1
gravestone	1
horse	1
house	45
human	1
hut	3
hydrogen atom	1
jenga	3
machine	1
marquis	1
minecraft steve	1
octagon	1
pyramid	1
room	1
sauna	1
school	1
shack	1
shed	2
ship	1
shop	1
sign	1
skyscraper	2
snowflake	1
space station	3
stable	3
stadium	2
star	1
T	1
tank	2
temple	6
toilet	1
tower	27
train station	1

training station	1
truck	1
tunnel	1
wall	3
zoo	1
N/A	6

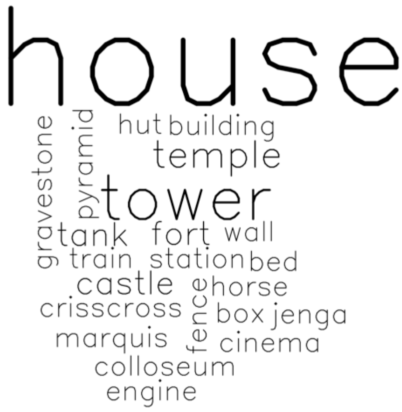


Figure 29: Answers to the question what children said they built in the social condition, n = 60

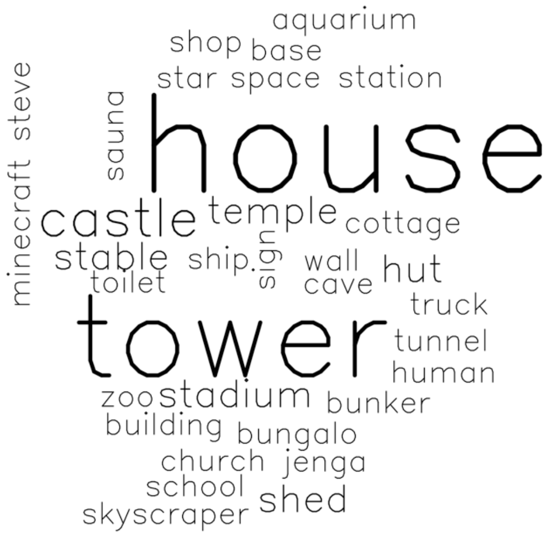


Figure 30: Answers to the question what children said they built in the asocial condition, n = 66

Figure 31: Answers to the question what children said they built in the asocial control condition, n = 40

4.3.1.3 *Details*

Individuals did not add more details to the description of what they built in any of the conditions. All models did not differ in their weights from the intercept-only model (Table 23; Figure 32) and the uncertainty was large.

Table 23: The WAIC values of models with the amount of detail included in the "what did you build" answers. M1.1 is the intercept-only model, m1.2 is the intercept, age, sex model and m1.3 is the intercept, age, sex and conditions model.

	WAIC	dWAIC	weight	SE	dSE
m1.1details	203.8	1.0	0.34	8.79	4.90
m1.2details	202.8	0	0.56	10.01	N/A
m1.3details	206.2	3.4	0.10	10.44	2.15

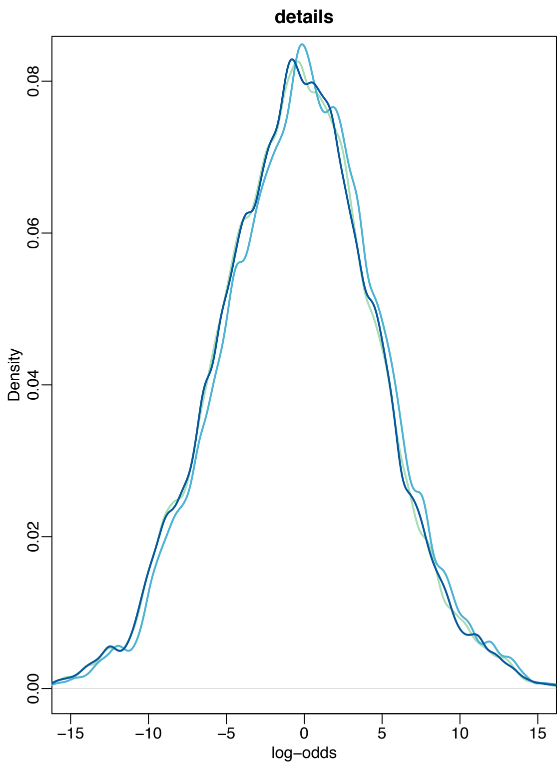


Figure 32: Log-odds of conditions from the models predicting the amount of detail added to the question "what did you build?" (m1.3 including age, sex and condition). Dark blue – social condition, green – asocial and light blue – asocial control.

4.3.2 Ratings of the pairs analysis

4.3.2.1 All conditions

The model best predicting the originality ratings of the pairs was the model with only the conditions as a predictor (Table 24).

Table 24: WAIC values of the models with response variable: rating of a pair of images. M1.1 included the sex category of the pair, average age of the pair and absolute difference in ages of the pair. M1.2 included the sex category of the pair, average age of the pair, absolute difference in ages of the pair and the conditions. M1.3 included the sex category of the pair, average age of the pair, absolute difference in ages of the pair, interaction between average ages of the pair and absolute difference in ages and the conditions. M1.4 included only the conditions as predictor variables.

	WAIC	dWAIC	weight	SE	dSE
m1.1ratings	399.2	8.8	0.01	21.14	6.58
m1.2ratings	396.7	6.3	0.04	22.56	1.29
m1.3ratings	398.5	8.1	0.02	22.65	1.84
m1.4ratings	390.4	0	0.93	22.42	N/A

The prediction for ratings lower than category 7 (very different) were in all conditions very uncertain. Pairs were less likely to be rated as 7 (very different) in the social and asocial conditions than in the asocial control condition (Figure 33). In the social and asocial condition, the ratings were similar (Figure 34).

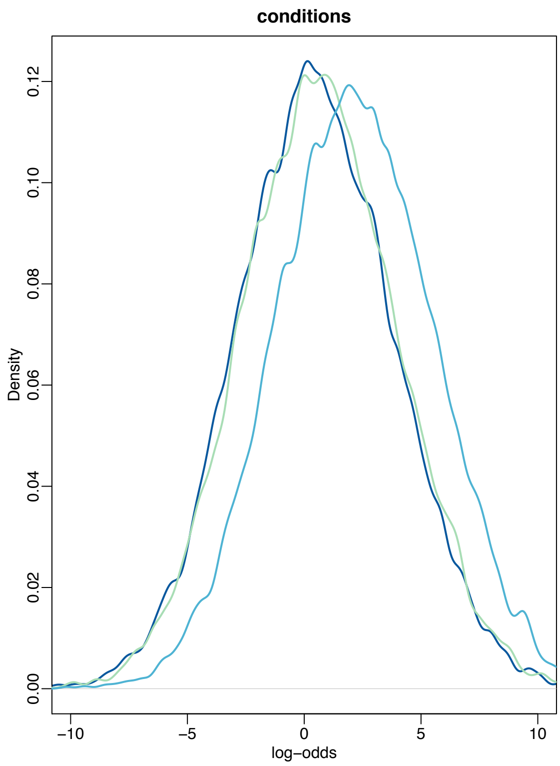


Figure 33: Log-odds of conditions predicting the similarity ratings of the pair from models including conditions as predictors. Dark blue – social condition, green – asocial and light blue – asocial control.

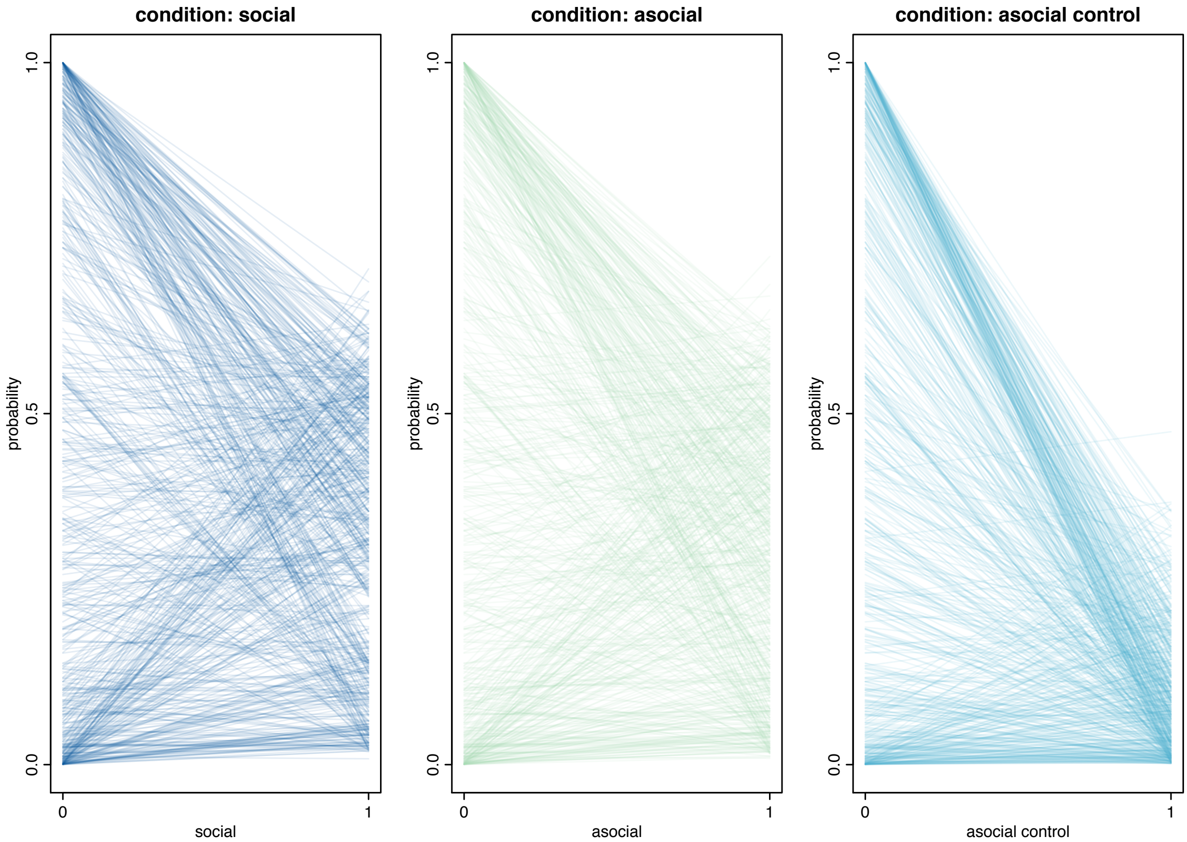


Figure 34: Posterior predictions of the ordered categorical models including conditions as predictors (m1.4). Each plot shows how the distribution of predicted response varies with condition: social, asocial, asocial control. Lines show where predicted thresholds between the ordinal categories lie in terms of cumulative probability. Distributions show cumulative probability, when they are stacked more towards the top of the graph, this shows the structures compared were more likely to be rated low on the Likert scale (as very similar), whereas when lines are stacked more towards the bottom, the cumulative probability is larger for higher values on the Likert scale (higher originality between structures compared).

4.3.2.2 Social and asocial – same time, same build, same model

In a comparison between the social and asocial condition the model that best predicted the data only included the behaviours I measured (Table 18). These were whether individuals in pairs built at the same time, shared the same model or were exposed to the same final build. However, no model had a Wakaike weight more than 0.9 suggesting that the prediction was highly uncertain (Table 25). Since I was interested in whether the different conditions children built in influenced their creativity, I used the model (m1.5) that included behaviours and conditions as predictors from which I assembled posterior prediction graphs (Figure 35, 36). In the asocial condition the partitions were opaque and the participants could not directly see each other. Whereas, in the social condition the partitions were transparent and they could see each other and other's building as well as final builds that might have been present at the IRP. There were small differences in these conditions (Figures 35, 36). In the social condition buildings that were built at the same time within the pair were slightly less likely to be rated as very different (Figure 35). The same pattern was observed when participants shared the same model who was building at another station at the IRP. In the asocial condition these patterns were not present. When participants in the asocial condition were exposed to the same build at the table it was less likely the pair was rated as very different and the rating predictions from the model did not have any clear pattern (Figure 36).

Table 25: WAIC values of the models from the social and asocial condition with response variable: rating of image pairs.

	WAIC	dWAIC	weight	SE	dSE
m1.1	359.2	7.3	0.02	17.38	5.86
m1.2	361.8	9.9	0.01	17.70	5.78
m1.3	362.6	10.7	0.00	17.32	3.32
m1.4	358.8	6.9	0.02	17.23	3.15
m1.5	351.9	0.0	0.79	16.85	N/A
m1.6	355.1	3.2	0.16	16.95	0.54

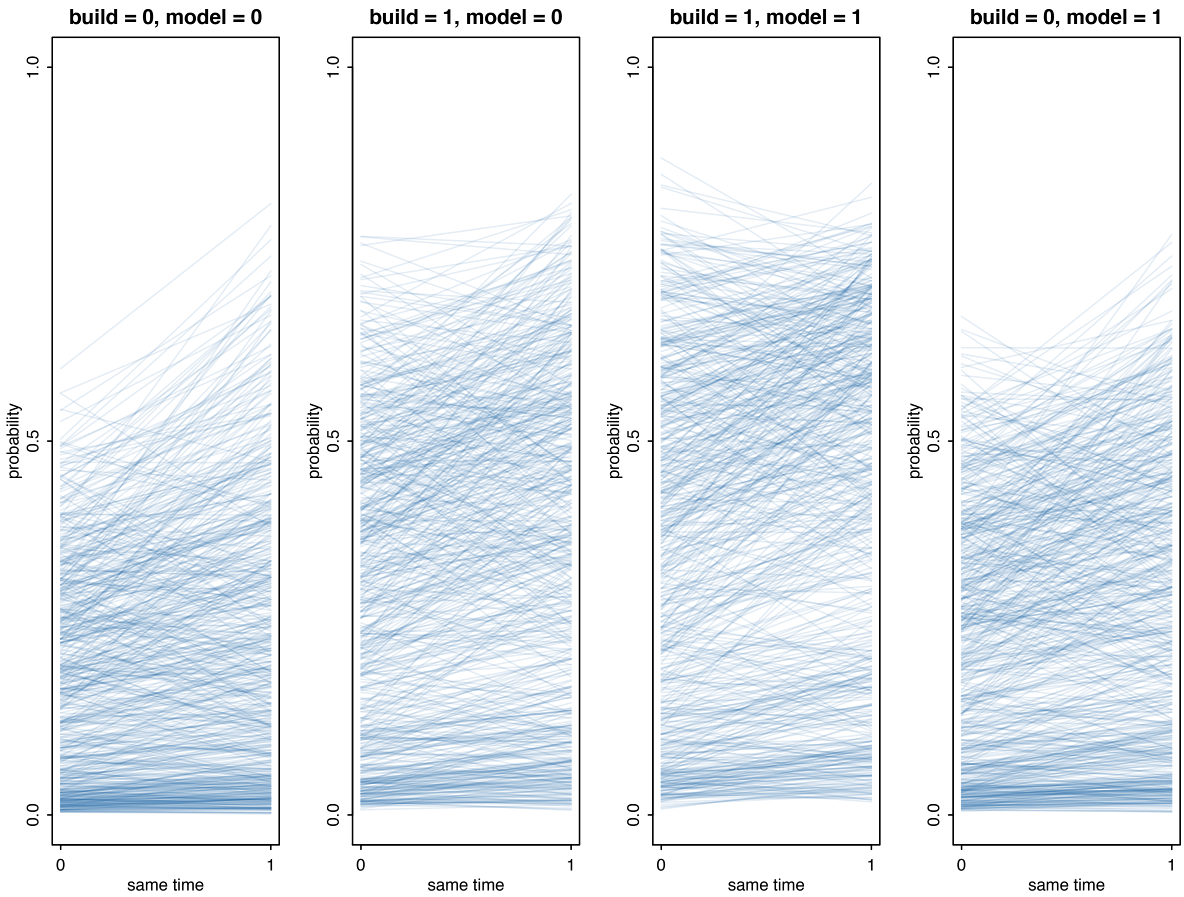


Figure 35: Posterior predictions of the ordered categorical models including conditions and behaviours as predictors (m1.5). The figure shows the distribution of predicted responses in the social condition. Each plot shows how the distribution of predicted responses varies when building at the same time at the Interactive Research Pod and when the pair were able to see the same final build (structure) or shared the same model who was building at another station. Lines show where predicted thresholds between the ordinal categories lie in terms of cumulative probability. Distributions show cumulative probability, when they are stacked more towards the top of the graph, this shows the structures compared were more likely to be rated low on the Likert scale (as very similar), whereas when lines are stacked more towards the bottom, the cumulative probability is larger for higher values on the Likert scale (higher originality between structures compared).

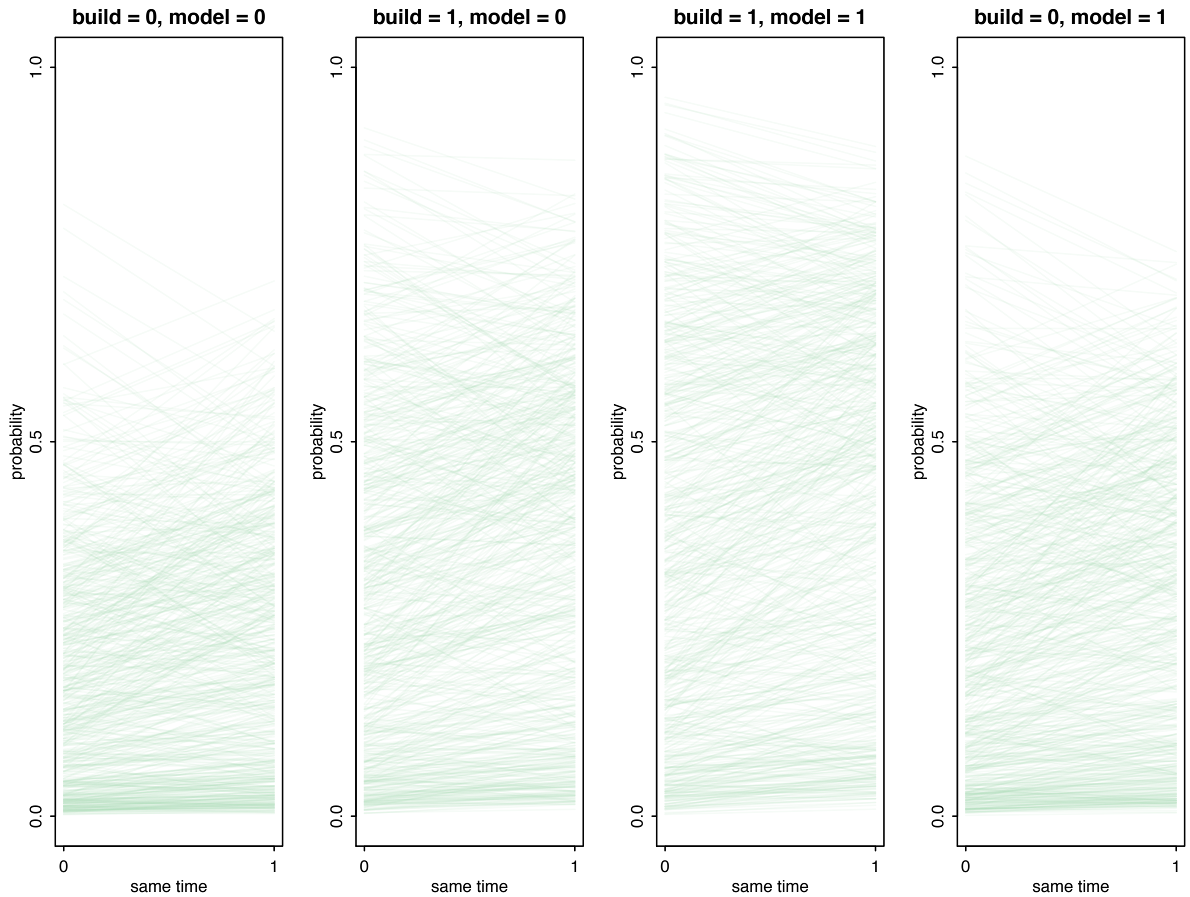


Figure 36: Posterior predictions of the ordered categorical models including conditions and behaviours as predictors (m1.5). The figure shows the distribution of predicted responses in the asocial condition. Each plot shows how the distribution of predicted responses varies when building at the same time at the Interactive Research Pod and when the pair were able to see the same final build (structure) or shared the same model who was building at another station. Lines show where predicted thresholds between the ordinal categories lie in terms of cumulative probability. Distributions show cumulative probability, when they are stacked more towards the top of the graph, this shows the structures compared were more likely to be rated low on the Likert scale (as very similar), whereas when lines are stacked more towards the bottom, the cumulative probability is larger for higher values on the Likert scale (higher originality between structures compared).

4.3.2.3 Social condition – behavioural effects and cluster effect

The pairs that were building at the same time were slightly less likely to be rated as very different (Figure 37, 38) in comparison to pairs that either shared a model that was building at another station or were exposed to the same final structure.

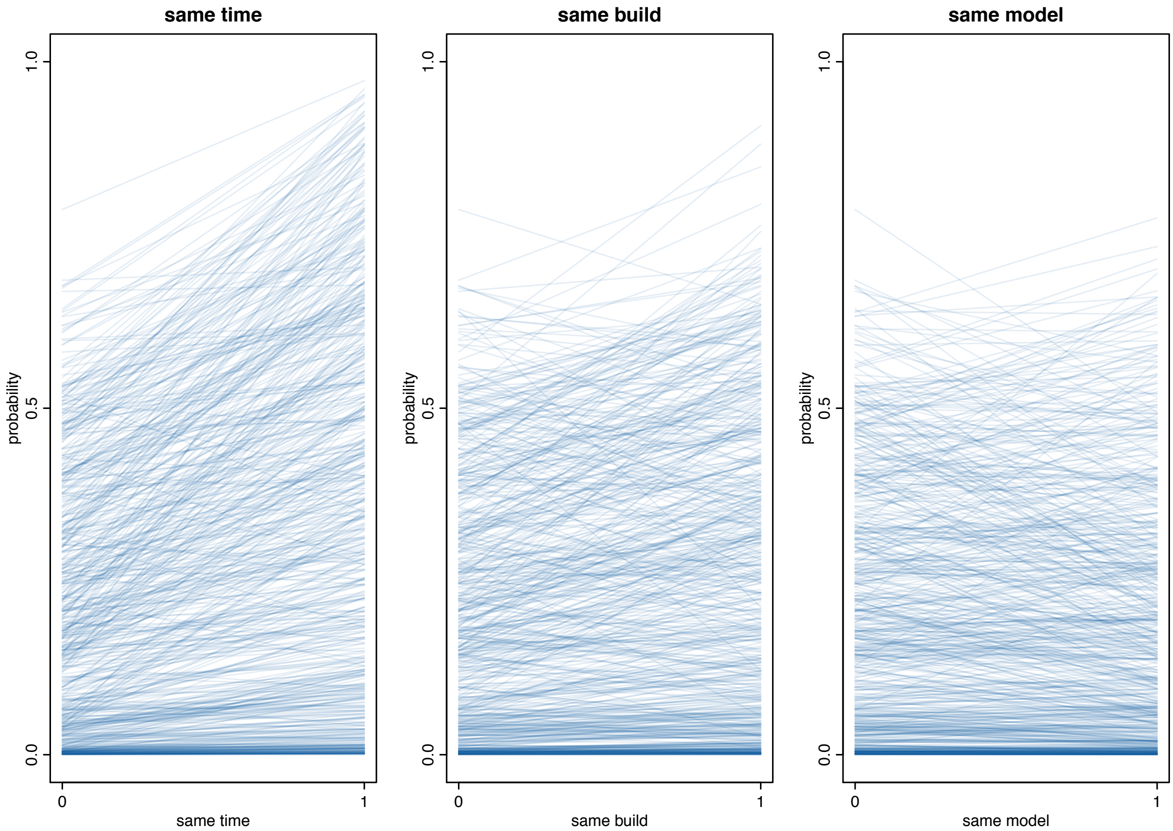


Figure 37: The posterior predictions of the ordered categorical models of participants building in the social condition including building at the same time, being exposed to the same final build or sharing the same model building at another table as predictors. Each plot shows how the distribution of predicted responses varies when building at the same time at the Interactive Research Pod and when the pair were able to see the same final build or shared the same model who was building at another station. Lines show where predicted thresholds between the ordinal categories lie in terms of cumulative probability. Distributions show cumulative probability, when they are stacked more towards the top of the graph, this shows the structures compared were more likely to be rated low on the Likert scale (as very similar), whereas when lines are stacked more towards the bottom, the cumulative probability is larger for higher values on the Likert scale (higher originality between structures compared).

Being exposed to the same build and sharing the same model did not affect the ratings pairs received (Figure 38).

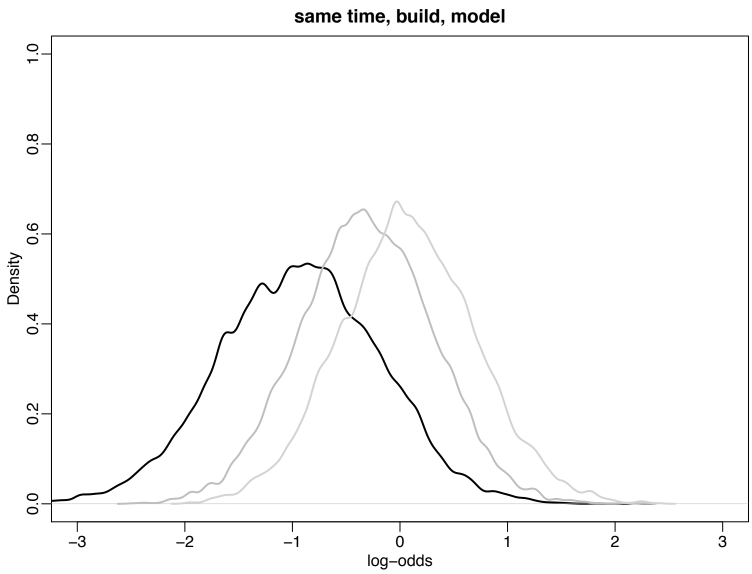


Figure 38: The log-odds of behaviours predicting the similarity ratings of the pair from the model including behaviours as predictors. Black – same time, grey – same build and light grey – same model.

When comparing models with predictors: building at the same time and being part of a cluster, building at the same time did not have a strong effect on the ratings that the pairs belonging to a cluster received (Figure 39). In contrast, the pairs that were building at the same time and not in clusters were not likely to be rated as very different.

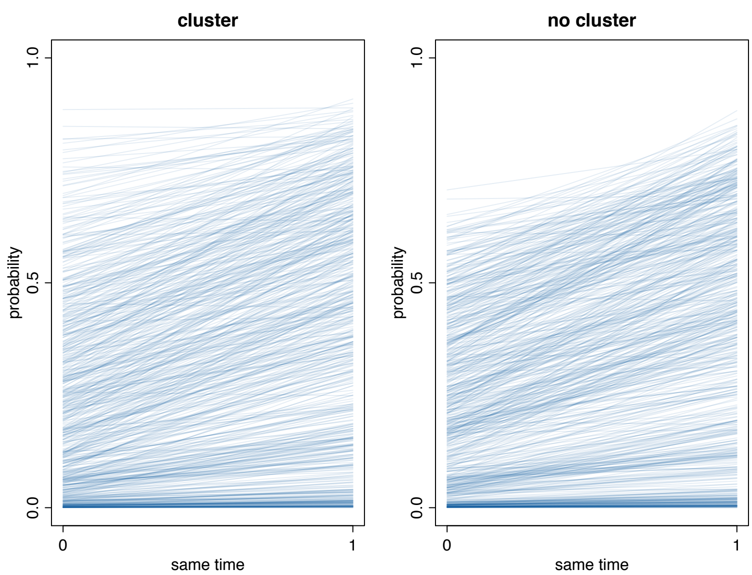


Figure 39: The posterior predictions of the ordered categorical models of participants building in the social condition including building at the same time and being part of a cluster. Each plot shows how the

distribution of predicted responses varies when building at the same time at the Interactive Research Pod in relation to being part of a bigger cluster. Lines show where predicted thresholds between the ordinal categories lie in terms of cumulative probability. Distributions show cumulative probability, when they are stacked more towards the top of the graph, this shows the structures compared were more likely to be rated low on the Likert scale (as very similar), whereas when lines are stacked more towards the bottom, the cumulative probability is larger for higher values on the Likert scale (higher originality between structures compared).

I ran a model with variable cluster as an intercept to see whether the eight clusters were rated similarly. There was no clear pattern within the clusters and some pairs in some clusters were mostly rated as very different and others as lower categories (1 – very similar to 7 – different) on the originality scale. The posterior means and 89% highest density intervals for clusters always cross zero. The HPDI shows that in some cases pairs in clusters were slightly more likely to receive lower ratings (g[2, 3, 8]) and in other cases higher ratings (g[1, 4, 5, 6]) (Figure 40).

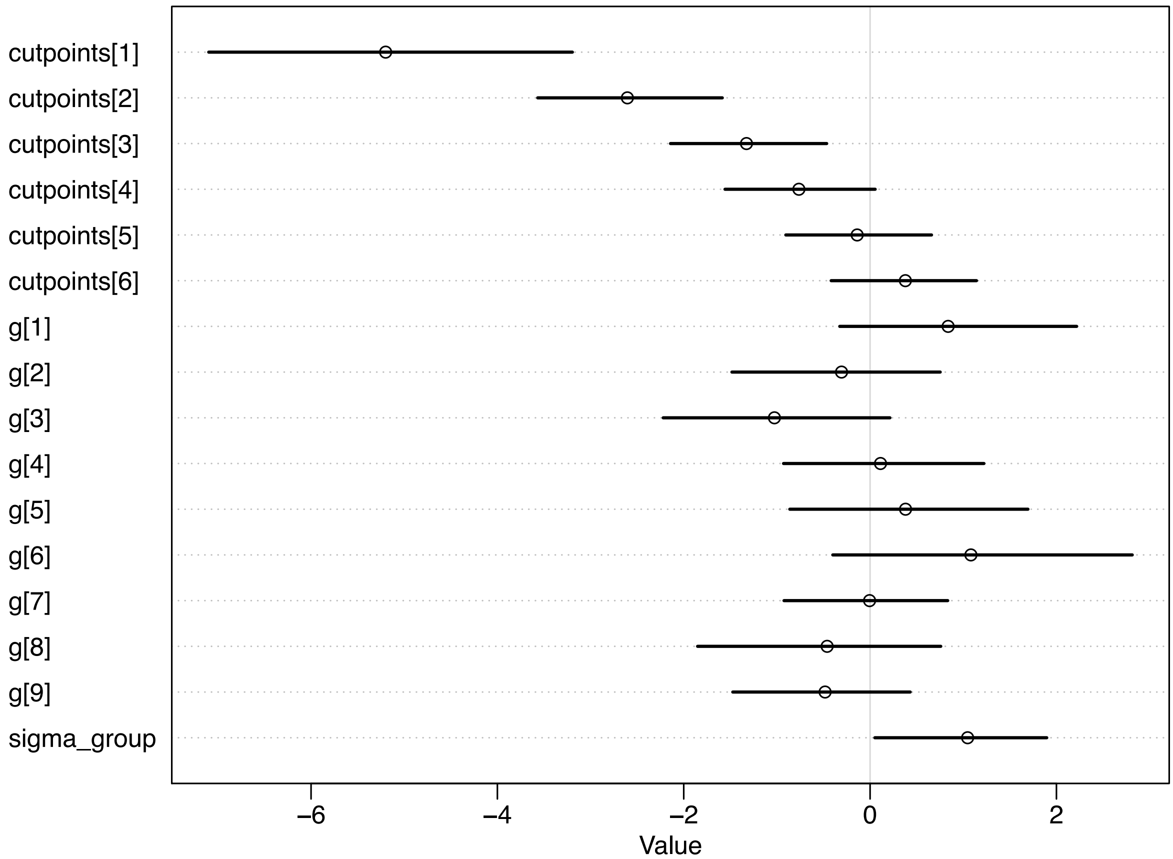


Figure 40: Posterior means of estimates and 89% highest posterior density intervals for the model including clusters as intercept. G[1-8] presents clusters and g[9] presents a cluster of pairs that did not belong to any of the clusters.

4.3.2.4 Heatmaps

Heatmaps show the variability of the ratings within the pairs of images in the same cluster. There is no trend among the clusters of pairs of structures being rated as similar (Figures 41, 42, 43, 44, 45, 46, 47, 48).

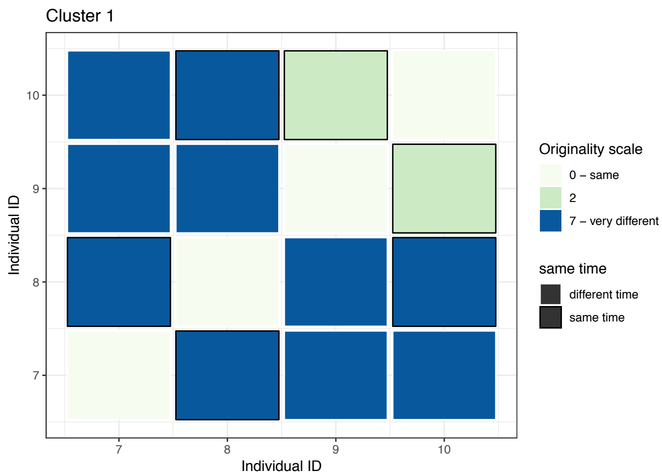


Figure 41: Cluster One with 4 individuals who gave **assent**, while 7 other[3] participants were building at the IRP at similar time. The black line around the squares represents the pairs that built at the same time.

[3] These participants either did not give assent, were not building alone, were adults or the recording was not suitable for analysis. The same is true for all other clusters.

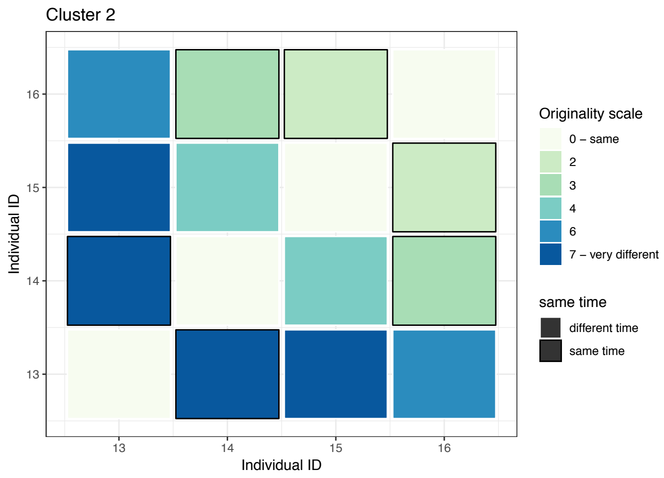


Figure 42: Cluster Two with 4 individuals who gave assent, while 5 other participants were building at the IRP at similar time. The black line around the squares represents the pairs that built at the same time.

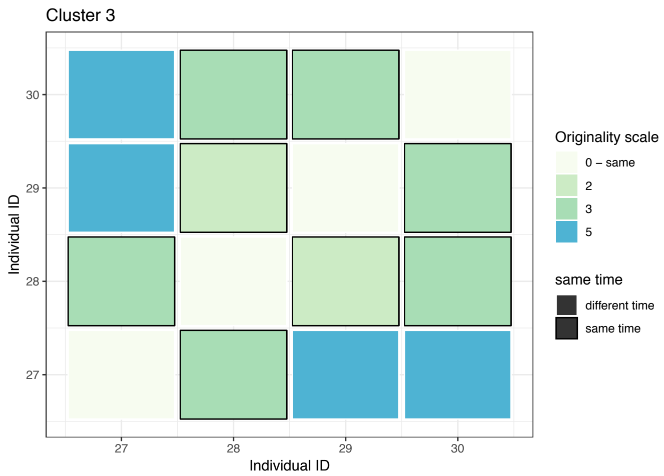


Figure 43: Cluster Three with 4 individuals who gave assent, while 4 other participants were building at the IRP at similar time. The black line around the squares represents the pairs that built at the same time.

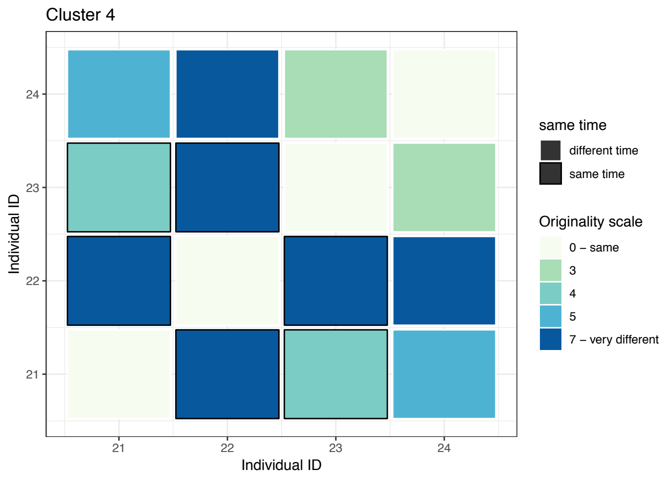


Figure 44: Cluster Four with 4 individuals who gave assent, while 13 other participants were building at the IRP at similar time. The black line around the squares represents the pairs that built at the same time.

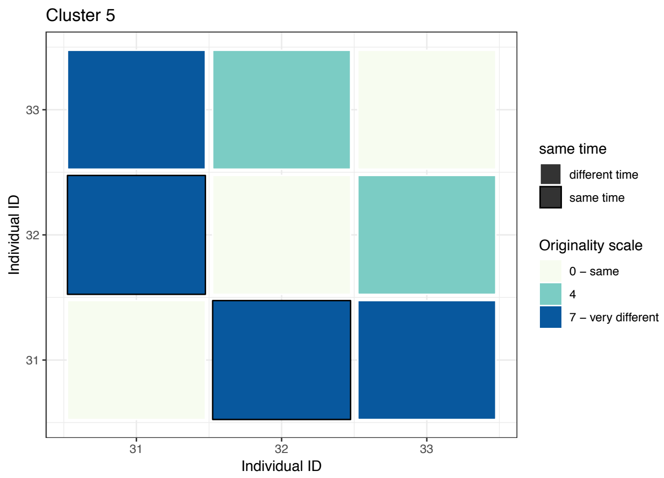


Figure 45: Cluster Five with 3 individuals who gave assent, while 12 other participant were building at the IRP at similar time. The black line around the squares represents the pairs that built at the same time.

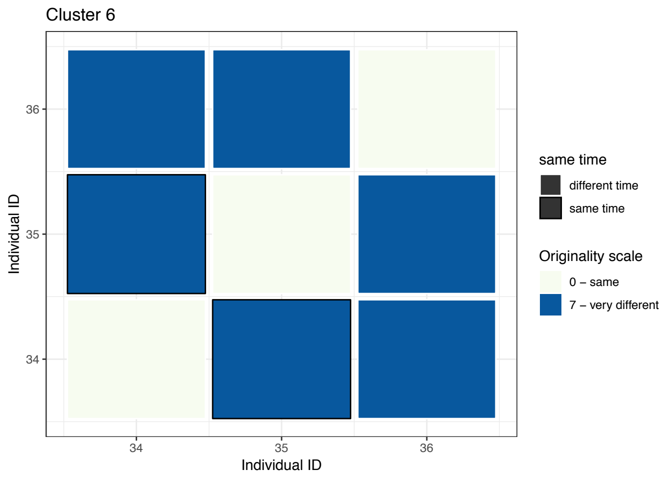


Figure 46: Cluster Six with 3 individuals who gave assent, while 13 other participants were building at the IRP at similar time. The black line around the squares represents the pairs that built at the same time.

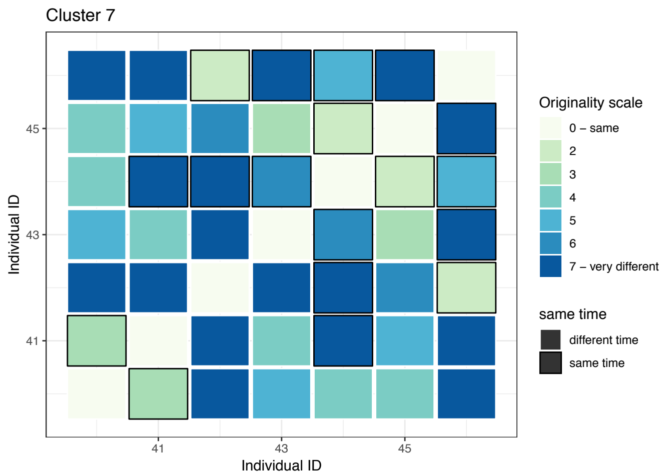


Figure 47: Cluster Seven with 7 individuals who gave assent, while 6 other participants were building at the IRP at similar time. The black line around the squares represents the pairs that built at the same time.

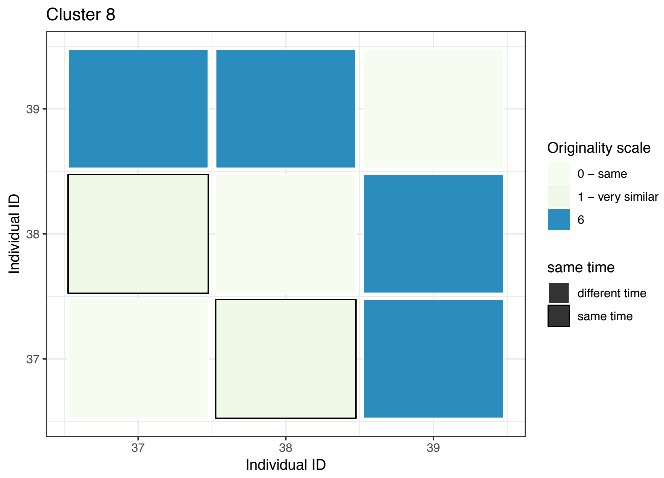


Figure 48: Cluster Eight with 3 individuals who gave assent, while 4 other participants were building at the IRP at similar time. The black line around the squares represents the pairs that built at the same time.

4.4 Discussion

4.4.1 What did children build in different conditions

Overall children described their structures in similar terms in all conditions (social – transparent partitions, asocial – opaque partitions, building at a similar time and asocial control – opaque partitions, building alone at the IRP and on different days). The most common structures were houses and towers. This might reflect the environment in which the children are embedded and what is deemed socially acceptable to build. Blocks or bricks are usually the foundations of houses and towers. This reveals an aspect of social learning not yet explored in an informal learning context. That is, the influence of norms when building and copying. Children might look for guidance on what is "correct", expected or socially acceptable (Burdett et al. 2016; Emma Flynn, Turner, and Giraldeau 2018; Göckeritz, Schmidt, and Tomasello 2014) to be built at the IRP. In this study we cannot know whether children copied structures or whether they built structures that were more common because of their predisposition to build familiar structures such as houses and towers.

4.4.2 Originality of the structures in different conditions

Overall the ratings adults assigned to pairs of structures in different conditions were high, suggesting that they were seen as being very different. The ratings of the pairs into other categories (very similar - rating 1, to very different - rating 7) were more uncertain in all three conditions. In this open-ended task children seemed to strive to build structures that were different to those of others. However, when comparing pairs between the social, asocial and asocial control conditions, there were differences in the probabilities of the pair being rated as very different. Pairs in the asocial control conditions were more likely to be rated as very different compared to pairs in both social and asocial conditions whereas the social and asocial conditions did not differ from each other. This partly corresponded with my predictions. Children that were building alone and whose structures were compared to structures of children who built on different days were most likely to be rated very different. Similar results were found when children built spaghetti towers in a transmission chain (Reindl and Tennie 2018). A reason for why children in the asocial control conditions built more original structures than those in the other conditions may be because in the asocial control condition children were truly building alone and thus did not have an opportunity to imitate any of the structures other participants were building on other stations by peeking around the opaque partitions. In this respect, the ‘asocial condition’ can be regarded as a social condition with less social information than the ‘social condition’.

The ratings of the pairs of structures in the social and asocial conditions were very similar. This may reflect the opportunity that children had to see others or their builds through transparent partitions or by peeking around the opaque partitions when in the asocial condition. This indicates that children do prefer social information when it is available and imitate even when it is unnecessary to do so (Turner, Giraldeau, and Flynn 2017). Previous studies found that children's over-imitation might be explained through normativity (Keupp, Behne and Rakocsy, 2015). Children that copy in the social and asocial conditions might consciously or subconsciously be attracted to information that is "conventional" (for example towers or houses) (McGraw et al. 2014). We cannot know whether the slightly higher probability for structures being similar in social and asocial condition stems from the fact that some types of builds have a predisposition or a higher probability of being assembled in comparison to other builds that are not assembled that often. However, since the towers and houses were also common in the asocial control condition, these two types of structures likely have a higher predisposition of being built in comparison to others, regardless of the condition children are in.

The largest impact on the ratings was seen when children were building at the same time, compared to when they could only see the final product (same build) or shared the same model at the IRP. This is in line with my prediction where I expected pairs of structures built by children building at the same time to be most likely to be rated as similar. First, when children were building at the same time, they were exposed to the process of building (micro-structures) and sometimes the final structures (macro-structures). Therefore, they might have copied the micro-structure by putting blocks together in the same way. Second, children are attracted to other peer's social cues

(Haun, Rekers, and Tomasello 2014; Köymen et al. 2014; Zmyj et al. 2012) and information from observing another child building might be more attractive than observing a built structure in the absence of the "builder". For younger children it might be easier to use the social information from the process of building than to copy final 3D structures due to the lack of spatial cognition skills which are still developing between ages 4 to 7 years (Casey et al. 2008). Third, children are known to copy the actions of a demonstrator (Whiten et al. 2009; Horner and Whiten 2005; Hopper 2010; Call, Carpenter, and Tomasello 2005; Tennie, Call, and Tomasello 2006) and they might have relied and copied the building processes of others building at the same time at other stations.

In the social condition I performed an additional analysis on the pair comparisons of the structures that children built in clusters. I identified a cluster as a group of children that either built at the same time, were exposed to the same final builds or shared the same model. There was no clear trend in comparing the pairs that were in different clusters with the pairs that did not belong to a cluster. Nor was there a trend among the clusters. In some clusters the pairs were more similar and in others they were completely different. These findings are not in line with previous studies of transmission chains in children where towers from the same chain are more similar to each other than towers from different chains (Reindl and Tennie 2018). The reason why structures in clusters were not necessarily very similar to each other may result from some individual differences between children's preferences to copy and rely on social information more than others (Mesoudi et al. 2016). Differences in copying preferences have been investigated by giving children an option to either select social information or start exploring the task on their own (Flynn, Turner, and Giraldeau 2016). It might be that some children's personalities lead to more copying or more inventiveness (Rawlings, Flynn, and Kendal 2017). I also cannot exclude any other factors that might influence whether they copy or not and who they copy in the specific context of the exhibit. The IRP enabled us to track whether any trends in buildings emerged but, in this study, I had a small sample size of clusters. In the future we could gather more cluster data from the videos from the IRP and perform more analysis regarding whether any traditions emerge in visitors building behaviours in this naturalistic context.

Since this study was performed in an informal learning context without experimental control I cannot claim that imitation occurred. Overall the results show that the structures children built in different conditions were different but not directly depending on the condition they built in which implies a limited role for social learning. Children and adults are prone to select social information if it is available (Turner, Giraldeau, and Flynn 2017), and they do so dependent on the expected difficulty of the task (Carr, Kendal, and Flynn 2015; Williamson, Meltzoff, and Markman 2008; Pinkham and Jaswal 2011). The task performed in this study was an open-ended building task that was not deemed challenging for children older than 4 years old, potentially explaining the lack of difference between social (social and asocial condition) and asocial (asocial control) conditions. The small differences among the three conditions, especially when participants built at the same time in the social condition, might be explained by conformity (Morgan, Laland and Harris, 2015; Flynn, Turner and Giraldeau, 2018) or in some cases by other biases when children knew each other in advance.

In the future the IRP could be used to test different types of social learning and social transmission among children and adults. For example, through employing the replacement method, used in social transmission studies, we could train a model (child or adult) to build an unstable structure and see whether others copied. In this case children might copy an inefficient structure over stable ones because the 'fun' element of an unusable structure might also be more appealing and provide some value to the structure. Therefore, providing insight into motivation and social learning biases in children.

The IRP would also enable us to study copying biases. By adjusting the questionnaire we could build a better picture about the relations between the members of the groups that build at the same time since children selectively imitate and learn from specific individuals (Wood, Kendal, and Flynn 2013b; Harris 2012). In a subsequent study I could explore in more detail some social learning strategies. It would be interesting to explore whether children copy others they know rather than strangers (Over and Carpenter 2012) and whether they prefer to copy adults or other peers (Wood, Kendal, and Flynn 2012; Hanna and Meltzoff 1993; Ryalls, Gul, and Ryalls 2000; Haun and Tomasello 2011) and whether they are more prone to copy groups or individuals (Evans et al. 2018) in the context of an informal learning environment. I might also want to explore biases to copy attractive or memorable builds, representing content biases. Furthermore, when more is known about the children that constitute a cluster, researchers can see whether any other biographic, social, cognitive and temperament predictors explain social learning well. I would expect that social learning biases connected to age, popularity, dominance, impulsivity and shyness are also present in this more natural context in groups of children.

4.4.3 Limitations

Since this study was performed “in the wild” (Flynn and Whiten 2010), without researchers being present and the data are messy, I was unable to assess any biases in regards to the model (age, sex, prestige, knowledge) thus I was not able to study social learning strategies (Rendell et al. 2011; Kendal et al. 2018; Wood, Kendal, and Flynn 2013b) children employed in natural settings. Furthermore, I cannot know whether children actually copied or not when building and I do not know the factors that led to the recorded differences between conditions.

4.5 Conclusion

This is a novel study of social learning and creativity in children in a naturalistic context. I focused on general principles of social learning and the impact on children’s creativity rather than the mechanisms behind it, such as transmission biases. In the future this type of setup with an adjusted questionnaire and the presence of researchers might enable us to also study all the social learning biases in the context of informal learning environment. These would inform academics working in the fields of education and social learning as well as museum practitioners who inform their visitors about the research being conducted.

Social learning is one of the pillars of human learning and is also present when we conduct science. Together with creativity it leads to scientific progress and the Interactive research pod enables people to experience a scientific discipline that is not usually presented in science centres, but is nonetheless very important for the development of human knowledge.

5 Cooperation in children and creativity

Cooperation and collaboration in pairs and groups of children fosters creativity

In this study a research assistant (Guy Lavender-Forsyth) helped me recruit the pairs and groups of children at the LFS. He will have co-authorship on any future publication(s) which might come out of this work.

Abstract

Children learn in groups and when they cooperate. Cooperation even fosters problem solving skills and creativity. Here I explored the direct connection between behaviours that are expressed when children build in groups, such as instructing, copying, discussing, cooperating and collaborating with the originality of the final builds children created in pairs and groups. I predicted that the extent to which children would cooperate and collaborate would be independent of their sex and age. I expected older children to instruct more than younger children, and younger children to copy more than older individuals. I predicted groups with an older average age would create more original structures than younger groups, and that cooperation and collaboration would lead to more original structures than instructing, adjusting others' builds and discussions. Furthermore, I predicted that structures would be less diverse and therefore less original in groups with a high incidence of copying than in groups where copying is low.

We recruited pairs (N=65) and groups (3 or more; N=28) of children (N=233) aged between 4 and 12 to build with the identical wooden blocks together at the Interactive Research Pod. They could build whatever they wanted. While they were building we coded the behaviours of interest and collected children's demographic data.

The behaviours, with the exception of instructions, were not influenced by the sex or age composition of the group or pair. Older children provided others with more instructions than did younger children, and these instructions slightly lowered the probability of structures being considered original. The extent of copying and discussing within pairs or groups did not have any influence on the originality of structures, whereas more cooperation and collaboration lead to more original structures.

Although cooperation and collaboration appeared to facilitate creativity, further research is required in more experimentally valid contexts to determine how strong this effect is.

5.1 Introduction

Humans imitate and teach from a young age (Tomasello 2009) and children from as young as 18-months old show pro-social and helping behaviours (F. Warneken and Tomasello 2006). By 4 years children provide instructions to others on how to solve tasks (Dean et al. 2012), understand shared goals (Carpenter, Akhtar, and Tomasello 1998; Meltzoff 1995; Gräfenhain et al. 2009) and they cooperate to solve problems researchers present them with (Ashley and Tomasello 2001; Warneken, Chen, and Tomasello 2006; Moll and Tomasello 2007). Besides understanding shared goals, altruistic behaviours and successful problem solving in groups, less is known about children's creativity and innovation in groups.

Previous studies that explored social learning in children followed dyads of children and used transmission chains (Flynn and Whiten 2008; Horner et al. 2006; Wood, Kendal, and Flynn 2013a) but did not investigate social learning and innovation in groups of children. Innovation in children is mostly studied on the individual level (Carr, Kendal, and Flynn 2015; Beck et al. 2011; Reindl et al. 2016; Nielsen 2013). For example, children in groups were more successful in retrieving rewards from puzzle boxes that required the accumulation of methods and techniques to progress compared to individuals solving puzzle boxes on their own (McGuigan et al. 2017). Studies using open diffusion methods to study groups of children and their solutions to a puzzle box focused on the methods of retrieving rewards and not on the processes children employed to solve the task (Flynn and Whiten 2012; McGuigan et al. 2017). When children are faced with a task in groups, they often show prosocial behaviours which include sharing rewards, cooperating with and teaching each other how to solve a puzzle (Dean et al. 2012; McGuigan et al. 2017; Hamann et al. 2011; Warneken et al.

2011). Furthermore, open diffusion experiments are usually performed in micro-societies in which groups of individuals are asked to solve a task. The tasks used in open diffusion experiments usually accommodate one participant at a time (see Dean et al. 2012 for an exception) which limits them to study behaviours of individuals rather than groups, leaving out an investigation into cooperation.

The sex of children might have an impact on their social learning (Frick, Clément, and Gruber 2017; Flynn and Whiten 2008; McGuigan and Cubillo 2013). Girls aged 3 to 5 years old are more likely to not finish a transmission chain experiment and are less successful in transmitting the information than boys (Flynn and Whiten 2008). However, studies using open diffusion did not record difference between sexes (Flynn and Whiten 2012). The sex of children does not influence their innovation skills (Carr, Kendal, and Flynn 2015). By the age of 7 and older children become as proficient innovators as adults (Beck et al. 2011; Cutting et al. 2014), but we can already detect creative skills in toddlers when they form jokes and express divergent thinking (Hoicka and Butcher 2016; Hoicka et al. 2018). Creativity is rarely tested in pairs and groups of children from 4 to 12 years of age. I therefore studied group innovation in children at a science centre. I asked children to build from wooden blocks in pre-formed pairs or groups. The task was selected as it does not limit children's creativity and is not a problem solving task which are commonly used in other studies of innovation in children (Beck et al. 2011; Nielsen 2013). To evaluate any differences among groups (according to their age and sex composition), I coded the relevant behaviours of interest. These were instructing others, adjusting the blocks placed by others in the structure, copying others, discussing, cooperating and collaborating.

In this study I wanted to explore whether the composition of groups and behaviours expressed within each group while building predicted creativity. I used an exhibit in a science centre where children could play and build from wooden blocks. With the help of a research assistant (GLF) I recruited pairs and groups of young participants. Children in pairs and groups could cooperate or work together on the same parts of the same structure or collaborate and work together on different parts of the same structure.

I expected that the extent to which children would cooperate and collaborate would be independent of their sex and age. Sex should not affect cooperation since children can freely decide who to cooperate with in the open setting of this experiment (Flynn and Whiten 2012). Age should not impact cooperation since children at very young ages express helping behaviours (Warneken and Tomasello 2006; Ashley and Tomasello 2001) and understand shared goals (Meltzoff 1995). I expected older children to instruct more than younger children due to them possessing more developed language skills (Ferrara et al. 2011), and younger children to copy more than older individuals as seen in previous studies (Flynn, Turner, and Giraldeau 2018; Clay, Over, and Tennie 2018; McGuigan et al. 2007; McGuigan, Makinson, and Whiten 2011). Based on the findings of studies investigating innovation in children where older children are better tool innovators than younger children (Beck et al. 2011), I predicted groups with an older average age would create more original structures than younger groups. As a previous study showed that children able to cooperate in a classroom developed better creativity skills than those unable to do so (Baloche 1994), I also predicted that cooperation and collaboration would lead to more original structures than instructing, adjusting and discussions. Alternatively, instructions provided by some individuals in the group could limit the space for novel ideas and therefore lead to less original structures than in groups where no instructions are articulated by group members (Bonawitz et al. 2011). Furthermore, I would predict that structures would be less diverse and therefore less original in groups with a high incidence of copying than in groups where copying is low (Reindl and Tennie 2018). However, it is possible that discussions among group members may lead to more original structures because the joint effort would result in a combination of diverse ideas (Muthukrishna and Henrich 2016).

5.2 Methods

The study was conducted in the Life Science Centre in Newcastle during the school summer holidays between the 15th of August to the 2nd of September 2017. I recruited 223 children (113 males; 4 to 13 years old) whom formed 93 pairs or groups that built together. For analyses, the pairs or groups were divided into categories based on sex (all male, all female, mixed), with varying numbers of participants in the groups (from 2 to 5) and relations among them (family, friends, did not know each other before the experiment) (Table 26). Consent was provided by the children's guardians (Appendix 3).

Table 26: The profile of pairs and groups

group size	relations in group	sex of individuals in group	range (years) of average group age	number of groups
2	family	females	6 – 10.5	13
		males	4.5 – 12	16
		mixed	5.5 – 12	28
	friends	females	8 – 10	4
		mixed	7.5 – 11	2
	not met before	females	9	1
		males	5.5	1

3	family	males	7.7	1
		mixed	7 – 11.7	10
	friends	males	10.7	1
	family friends	females	8 – 9	2
		mixed	6.7 – 9	4
	not met before	females	11	1
		mixed	8.7	1
4	family	males	11.5	1
		mixed	7.3 – 9	3
	family friends	mixed	8 – 10.8	3
5	family friends	mixed	11.4	1

5.2.1 Apparatus

The Interactive research pod (IRP) exhibit is in the Brain Zone of the Life Science Centre, which was developed through cooperation between centre staff and academics from Durham University (Kendal et al., 2016, Rudman et al., 2017). The IRP exhibit comprised of a triangular desk. On the desk were 100 Keva planks. These were cuboid wooden blocks, which were all the same brown colour and the same size 6.35 mm thick, 19.05 mm wide and 114.3 mm long.

5.2.2 Procedure

Participants were first introduced to each other in the case when they did not know each other. We asked them about their age and how many 'best friends' they had. The latter was an approximation of their extroversion (Watson and Clark 1997) and we used the number of responses they gave to approximate the size of their close friendship circle. For analyses, these answers were divided into two groups: children with less than 5 friends and children with 5 and more friends. After the introductions we requested each group to build something together from the Keva planks. Each group received the same instruction: "Here are the wooden blocks. You can build whatever you want, and you have as much time as you want. Try to build together with others in a group." No further instructions were provided. We did not limit their time building. When they said they were finished, the researchers asked what they built and they received a reward of a sticker for their participation in the experiment.

5.2.3 Behavioural measurements

The researchers (ZM and GLF) coded behaviours while individuals were building. The behaviours were divided into two groups: individual behaviours and group behaviours. Detailed explanations of the behaviours can be found in Table 27. Behaviour counts were adjusted for the time building. Group behaviours were counted at the group level, whereas individual behaviours were counted for each participant separately.

Table 27: Ethogram of behaviours measured at the IRP

Individual Behaviours	
Instructing	one participant tells or demonstrates to another one how to build a structure (e. g. "put blocks together like I do", "put the block there", "look, put the block like I do" (at the same time the child demonstrates), etc.)
Adjusting	one individual significantly changes the position of at least one block that another participant had placed into the construction (changes the direction, position, placement, orientation)
Copying Separately	participants are not contributing to the same structure and when a participant starts building their own build (structure) but follows the previously seen build (structure) assembled by another participant or when s/he changes the way they are building their own structure to match that of their partner(s)
Copying Same	when two or more participants are building the same structure and one participant starts building the structure in the same way as another participant and s/he is not being instructed or lead by another participant (e.g. we noticed the gaze of one participant directed towards the way another is putting blocks together or we saw that the way s/he is putting blocks

	together exactly mimics the way another participant was putting blocks together)
Resource Provision	when a participant collects blocks that are out of reach of another individual and gives them to him/her to build with; they can give the blocks directly into the other participants hand or just in the vicinity of the other participant, where s/he can reach them, and at the same time the participant providing the blocks does not use them
Group Behaviours	
Separation	when participants start building together but then, start building their own builds or are prompted to build on structure together by a researcher and start with their own builds
Cooperation	when at least two participants are building together and contributing to the same structure at the same time (e. g. are not talking with each other but are building the same structure together and the same parts of the structure, within the same 10 cm2 area of the structure)
Collaboration	when at least two participants are building the same build but different parts of it (e. g. are not talking with each other but are only building the same structure together but not the same parts of the structure)
Discussion	when at least two participants talk about what to build, both contributing suggestions (e. g. participants use questions, like “what do you want to build?”, “how do you want to build it?”, “do you think it would work …?”)

5.2.4 Ratings of the pairs of structures built by different groups of children

To objectively grade the originality of the structures built by different groups of children, I recruited adult participants through Prolific to rate pairs of structures based on their similarities. I developed a web application in the java script programming language using jsPsych library (de Leeuw, 2015) that presented randomly assigned pairs of wooden structures to adult raters. They were asked to rate them based on how structurally similar they appeared to each other on a scale of 1 (very similar) to 7 (very different). Each adult rated 110 pairs. I recruited 74 participants (36 males, median age: 28 years). Each rater received compensation of £1.25.

5.2.5 Statistical analysis

I divided the statistical analysis in two parts. First, I focused on the behaviours of individuals in groups and explored within-group individual variation. Second, I focused on between-group variation focusing on group behaviours and the comparison of the originality scores of the structures that the groups of children built together.

For the analysis of the individual behaviour, I counted the number of times an individual provided instructions to other group members, adjusted the position of the blocks that others placed in the structure, copied when building the same structure, copied when building separate structures and when an individual provided resources for others (Table 27). The last two listed behaviours were very rarely expressed. The low count of behaviours led to a floor effect (Martin and Bateson 2015) so I did not use copying separately and resource provision in any of the analyses.

I used the counts of three behaviours as response variables: instructing, adjusting and copying while building the same structure. Since the distribution of the response variables looked to be zero-inflated I used zero Poisson models in which one linear model predicted the probability of the behaviour count being zero and the other linear model predicted the average number of counts of behaviours when I took specific parameters into account. The 'null' model comprised of the intercept and the log of the time pairs or groups were building. The subsequent models included parameters for sex, age, the number of best friends individuals said s/he had, and group ID (Table 28) as predictor variables. I used weakly informative priors.

Table 28: Models with parameters predicting the counts of behaviours of an individual (instructing, adjusting and copying)

model	parameters included in the model
m1.1	time spent building
m1.2	time spent building + group ID
m1.3	time spent building + group ID + sex
m1.4	time spent building + group ID + age
m1.5	time spent building + group ID + sex + age
m1.6	time spent building + group ID + number of best friends
m1.7	time spent building + group ID + sex + age + sex*age
m1.8	time spent building + group ID + sex + age + sex*age + number of best friends

In the second part of the analysis I focused on group behaviours and the originality of the structures children built together in groups. First, I analysed the behaviour data between groups and second, the ratings of how original the structures were. The group behaviours I counted were cooperating, collaborating and discussing. Cooperating and discussing were present in a lot of the groups and were not rare, therefore I modelled them as response variables in models using a Poisson distribution. Collaborating was rarely seen so I used a zero Poisson distribution in the models. Groups that did not show any of the three behaviours were scarce. The response variables in models were the behaviour of interest: the number of times children cooperated, discussed and collaborated. The 'null' models in all cases were the intercepts and the log of the time each group was building for. Other parameters of interest included the sex of participants in the group and their average age (Table 29). I used weakly informative priors.

Table 29: Models with parameters predicting the counts of behaviours in groups (instructing, discussing, cooperation, collaboration and copying)

model	parameters included in the model
m1.1	time spent building
m1.2	time spent building + sexes of the participants in group

m1.3	time spent building + average age of the group
m1.4	time spent building + average age of the group + sexes of the participants in group

Lastly, I analysed the originality (composed of the mode of similarity ratings between pairs of structures) of what children in groups built. I explored whether group behaviour counts (cooperation and collaboration), and the total number of times the individuals in a group instructed and copied, predicted the originality score. I focused on five behaviours: instructing, discussing, cooperating, collaborating and copying (the sum of copying the same structure and copying a separate structure; see Table 27). I used an ordered logistic regression where the intercepts were all the categories on the Likert scale and the slopes were counts of specific behaviours. To ease the interpretation of results and not to overfit the models I modelled each behaviour and its relation to originality scores separately. I used weakly informative priors.

All statistical analyses were carried out in R 3. 3. 2. (R Core Team 2013) using the *rethinking* package (McElreath 2016). Posterior estimates were generated using RStan 2.17.3 (Stan Development Team 2018). Inter-coder reliability with Cohen's Kappa coefficient has been calculated using the *irr* package (Gamer, Lemon, and Singh 2019) in R.

5.2.5.1 Inter-coder reliability

To evaluate the reliability of the coders an additional 25% of the randomly chosen behavioural data were coded by the research assistant (GLF). Two researchers coded the behaviours in pairs and groups of children simultaneously, for all variables (counts of instructing, adjusting, copying, discussing, cooperating and collaborating). A Cohen's kappa of .896 was achieved, demonstrating high reliability.

As an additional test following Bayesian statistics methods to evaluate the counts I compared WAIC values and their weights between an intercept only model and a model including coder ID as a predictor. Models used either Poisson distribution or zero-Poisson distribution, had weakly informative priors and were approximated by a Gaussian distribution with 'map' function in *rethinking* package (McElreath 2016). There were no differences between the intercept only and coder ID models (Table 30). Therefore the coding was reliable.

Table 30: Models of inter-coder reliability.

behaviour	model	WAIC	pWAIC	dWAIC	weight	SE	dSE
instructing	mIntercept	349.5	2.1	0	0.73	21.27	N/A
	mCoder	351.5	3.3	2	0.27	21.50	1.12
adjusting	mIntercept	286.8	1.8	0.4	0.45	19.05	N/A
	mCoder	286.4	2.6	0.0	0.55	19.02	2.38
copySame	mIntercept	344.3	1.7	0.0	0.74	15.83	N/A
	mCoder	346.3	2.8	2.1	0.26	16.20	0.87
copying	mIntercept	374.2	1.5	0.0	0.72	12.68	N/A
	mCoder	376.0	2.3	1.9	0.28	12.96	0.39
cooperation	mIntercept	180.6	0.7	0.0	0.71	10.94	N/A
	mCoder	182.4	1.5	1.8	0.29	11.24	0.47
collaboration	mIntercept	103.9	1.4	0.0	0.7	13.76	N/A
	mCoder	105.6	2.0	1.7	0.3	14.17	0.42
disscussion	mIntercept	154.7	0.8	0.0	0.7	7.59	N/A
	mCoder	156.3	1.7	1.7	0.3	7.71	0.14

5.3 Results

5.3.1 Individual's behaviours

The behaviours of interest (response variables) were; providing instructions, adjusting blocks that others in the group placed in the structure and copying others that were building the same structure. None of the listed behaviours were well explained with the parameters sex, age and the number of best friends. The model weights and WAIC values did not differ between models (Table 31).

Table 31: WAIC values and Akaike weights from models with parameters predicting the counts of behaviours of an individual (instructing, adjusting and copying). Parameters included in the models are listed in Table 25.

model	WAIC	pWAIC	dWAIC	weight	SE	dSE
m1.6ins	599.84	28.59	0.00	0.39	32.81	NA
m1.8ins	600.15	27.82	0.30	0.34	32.55	1.92
m1.4ins	601.15	30.26	1.31	0.20	34.03	3.98
m1.5ins	603.49	26.93	3.65	0.06	33.94	3.04
m1.7ins	612.57	27.68	12.73	0.00	32.83	9.61
m1.2ins	615.05	25.79	15.20	0.00	32.92	9.15
m1.3ins	616.11	26.14	16.27	0.00	33.03	9.57
m1.1ins	623.92	3.21	24.08	0.00	34.75	12.27
m1.4adj	565.24	27.48	0.00	0.56	28.15	NA
m1.5adj	566.70	27.90	1.46	0.27	28.24	0.56
m1.6adj	568.65	28.68	3.42	0.10	28.45	0.92
m1.3adj	571.17	27.17	5.93	0.03	29.18	3.77
m1.2adj	571.69	27.48	6.46	0.02	29.37	4.23
m1.8adj	573.21	28.00	7.97	0.01	29.15	5.00
m1.7adj	576.53	26.05	11.29	0.00	30.08	5.47
m1.1adj	586.44	3.79	21.21	0.00	34.31	14.06
m1.1CopySame	530.97	3.28	0.00	0.21	30.03	NA
m1.3CopySame	531.19	19.27	0.22	0.19	30.13	3.47
m1.4CopySame	531.34	15.80	0.37	0.17	30.04	3.63
m1.2CopySame	531.47	17.31	0.49	0.16	30.22	2.47
m1.5CopySame	531.79	17.94	0.81	0.14	30.00	4.37
m1.6CopySame	532.82	16.83	1.85	0.08	30.19	4.51
m1.7CopySame	534.24	13.78	3.27	0.04	30.85	3.53
m1.8CopySame	536.86	17.31	5.89	0.01	30.79	5.46

5.3.2 Group behaviours

Group behaviours, cooperation and collaboration, did not differ depending on the average age of group members and whether they were male, female or mixed sex (Table 32).

Table 32: WAIC values and Akaike weights from models with parameters predicting the counts of behaviours of groups (cooperation, collaboration and discussion). Parameters included in the models are listed in Table 26.

model	WAIC	pWAIC	dWAIC	weight	SE	dSE
m1.1GCoop	281.56	0.91	0.00	0.34	9.96	NA
m1.2GCoop	282.30	1.76	0.74	0.24	9.74	1.11
m1.3GCoop	282.40	1.54	0.84	0.22	10.12	1.45
m1.4GCoop	282.64	2.42	1.08	0.20	9.83	2.32
m1.1GColla	174.16	2.19	0.00	0.43	18.70	NA
m1.2GColla	174.95	3.22	0.79	0.29	18.62	1.46
m1.3GColla	175.81	2.94	1.65	0.19	18.93	0.56
m1.4GColla	177.24	4.39	3.08	0.09	18.86	2.04

The behaviours, instructing and copying other individuals, did not differ between groups based on 'sex' (Table 33).

Table 33: WAIC values and Akaike weights from models with parameters predicting the counts of behaviours of groups (instructing and copying). Parameters included in the models are listed in Table 26.

model	WAIC	pWAIC	dWAIC	weight	SE	dSE
m1.3GIns	282.64	1.62	0.00	**1.00**	10.16	NA
m1.4GIns	368.12	8.51	85.48	0.00	23.67	22.77
m1.1GIns	378.64	4.02	96.01	0.00	28.94	27.54
m1.2GIns	381.70	7.22	99.06	0.00	28.84	27.43
m1.3GCopy	343.36	5.71	0.00	**0.74**	18.60	NA
m1.4GCopy	345.71	7.91	2.35	**0.23**	18.80	2.36
m1.1GCopy	350.52	3.82	7.16	0.02	19.67	9.19
m1.2GCopy	353.00	5.78	9.63	0.01	19.70	9.39

Copying increased in groups where the average age of children was higher (Figure 49). The models indicate that copying did not only depend on the average age of the children in the groups, but on the interaction between sex and average age of children in the group (Table 33). Although, the uncertainty of the models is high.

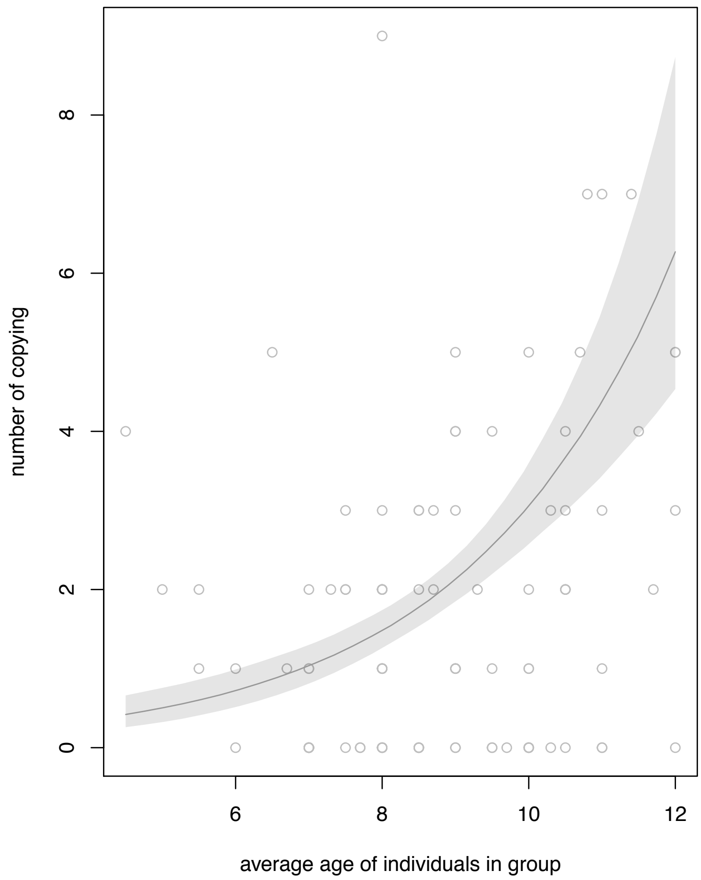


Figure 49: Posterior predictions together with actual measured data. Posterior predictions are drawn from m1.3GCopy. The blue dots are the data points from the actual data collected. The blue line presents the median number of copying given the average age (in years) of individuals in the group, the shaded area presents 89% interval.

Instructions were more often given in groups where the average age of children was higher than in groups that were on average comprised of younger children (Figure 50).

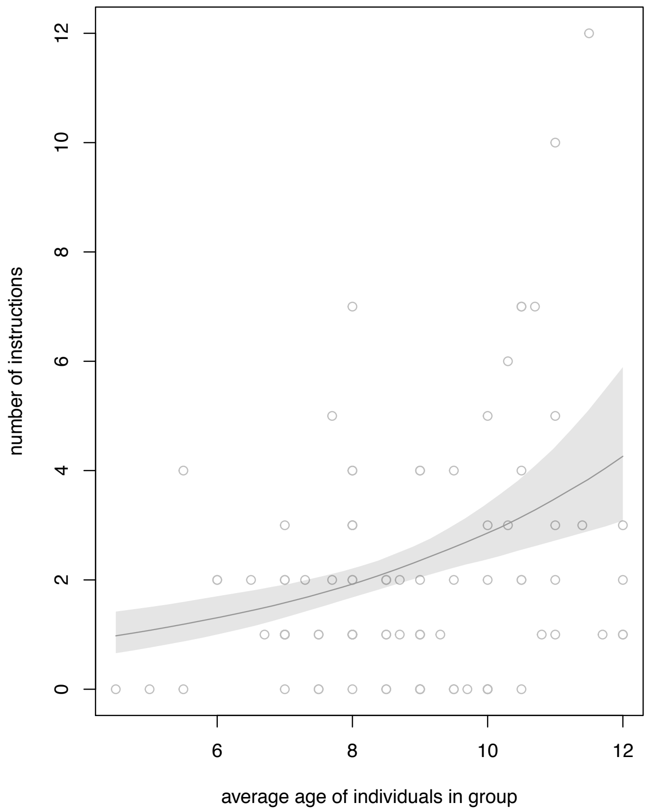


Figure 50: Posterior predictions together with actual measured data. Posterior predictions are drawn from m1.3GIns. The blue dots are the data points from the actual data collected. The blue line presents the median number of instructions given the average age (in years) of individuals in the group, the shaded area presents 89% interval.

5.3.3 "What did you build?" Answers

In this study the children in pairs and groups together decided what they assigned structure to be. Most common were houses and towers (Table 34).

Table 34: Count of the answers to what children said they built

"What did you build?"	Count
bunny rabbit	1
building	3
elephant	1
temple	3
railway track	3
stage	1
toilet	1
well	1

camp side	1
face	1
human	2
hospital	1
robot	1
stairs	1
tower	13
apartments	1
carwash	1
flats	2
hut	2
Metro logo	1
shed	1
star	1
furniture	1
castle	9
hexagon	1
town	1
skyscraper	1
wall	1
dominoes	1
house	21
jenga	3
plane	1
snake	1
Life Science Centre logo	1
box	1
ladybug	1
pyramid	1
snowman	1
theatre	1
don't know	3

5.3.4 Group behaviours and originality of builds

In the analysis I used the parameter counts of five behaviours as predictors of the ratings of the pair of structures children built. More instructions lead to structures in the pair being less likely rated as very different than when there were fewer instructions in groups (Figure 51).

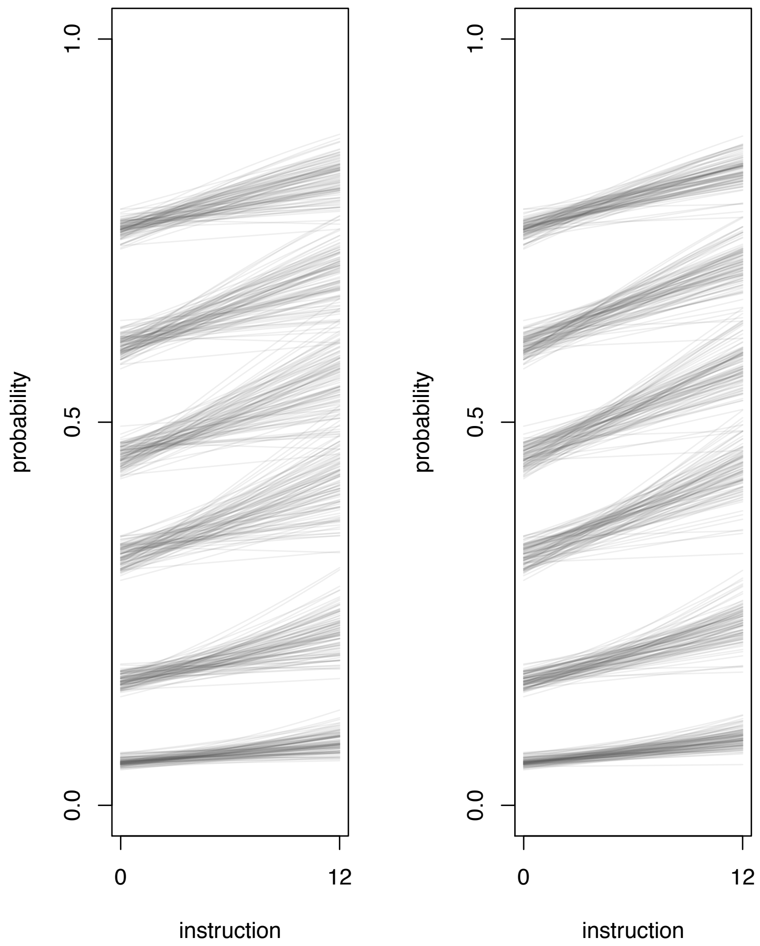


Figure 51: The posterior predictions of the ordered categorical models of participants building in pairs or groups in which there were different counts of instructions given (range 0 - 12). Each plot shows how the posterior probability distribution of predicted responses (ratings on the scale from 1 - very similar to 7 – very different) varies when more instructions were given by group members at the Interactive Research Pod. Since the raters rated two images at the same time, the left and right graph correspond to the pair being compared and the amount of instruction in either the left or right image in the pair. For example, one image in the pair is set up to 0 counts of instructions while the other image in the can have a range between 0 to 12. Lines show where predicted thresholds between the ordinal categories lie in terms of cumulative probability. Distributions show cumulative probability, when they are stacked more towards the top of the graph, this shows the structures compared were more likely to be rated low on the Likert scale (as very similar), whereas when lines are stacked more towards the bottom, the cumulative probability is larger for higher values on the Likert scale (higher originality between structures compared).

The presence of copying of the building process within the same group members did not change the probability of the structures being rated as more or less similar to other structures pairs (Figure 52).

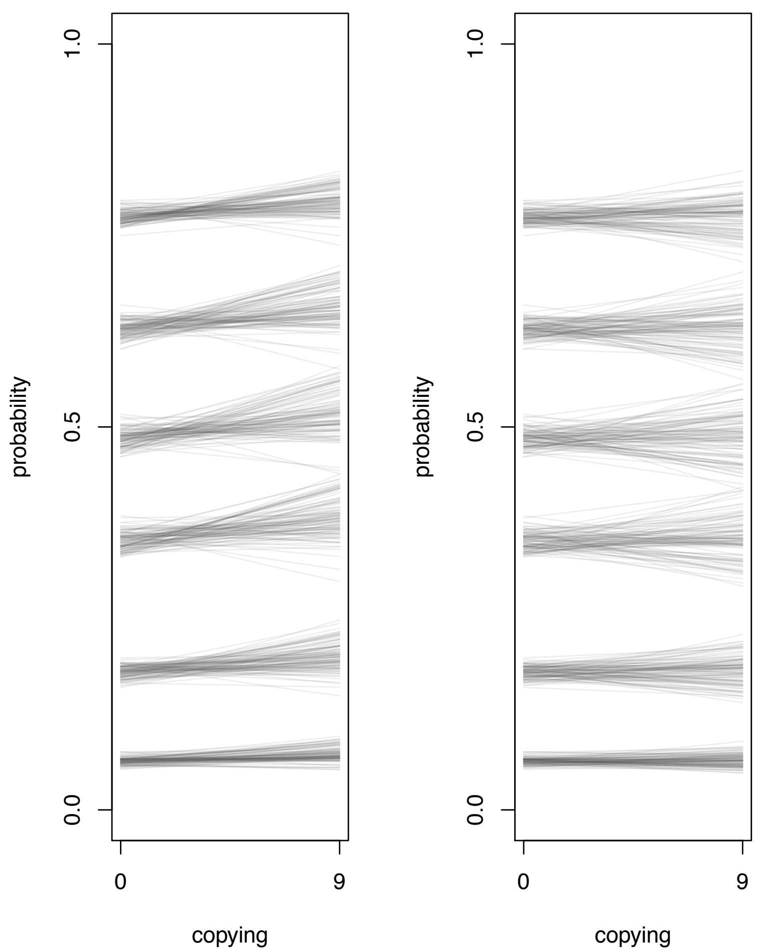


Figure 52: The posterior predictions of the ordered categorical models of participants building in pairs or groups in which there were different counts of copying (range 0 - 9) recorded. Each plot shows how the posterior probability distribution of predicted responses (ratings on the scale from 1 - very similar to 7 – very different) varies when more copying happened within the pair or group members at the Interactive Research Pod in the cooperative condition. Since the raters rated two images at the same time, the left and right graph correspond to the pair being compared and the amount of copying in either the left or right image in the pair. For example, one image in the pair is set up to 0 counts of copying while the other image in the can have a range between 0 to 9. Lines show where predicted thresholds between the ordinal categories lie in terms of cumulative probability. Distributions show cumulative probability, when they are stacked more towards the top of the graph, this shows the structures compared were more likely to be rated low on the Likert scale (as very similar), whereas when lines are stacked more towards the bottom, the cumulative probability is larger for higher values on the Likert scale (higher originality between structures compared).

When participants cooperated or built together at the same parts of the structure, the ratings of the structures were slightly more likely to be rated as different, meaning the structures were more original (Figure 53). Caution is needed with interpretation, however, since the higher the counts of cooperation, the more uncertain the ratings.

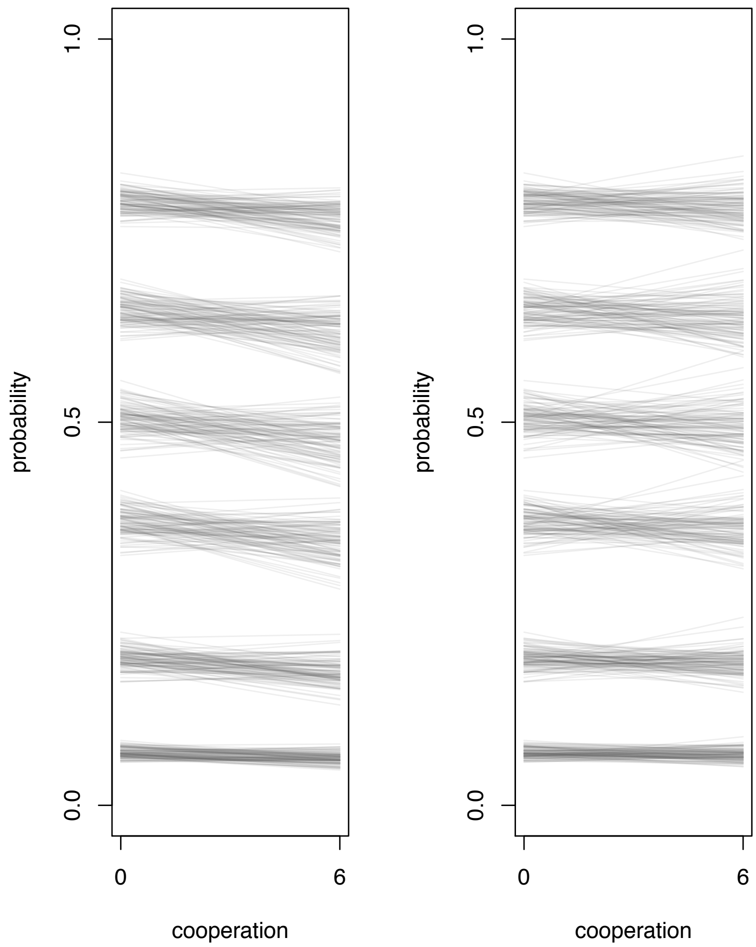


Figure 53: The posterior predictions of the ordered categorical models of participants building in pairs or groups in which there were a different counts of events of cooperation (range 0 – 6). Each plot shows how the posterior probability distribution of predicted responses (ratings on the scale from 1 - very similar to 7 – very different) varies when more cooperation occurred in the pair or group members at the Interactive Research Pod in the cooperative condition. Since the raters rated two images at the same time, the left and right graph correspond to the pair being compared and the amount of cooperation that occurred when children built either the left or right image in the pair. For example, one image in the pair is set up to 0 counts of cooperation while the other image in the can have a range between 0 to 6. Lines show where predicted thresholds between the ordinal categories lie in terms of cumulative probability. Distributions show cumulative probability, when they are stacked more towards the top of the graph, this shows the structures compared were more likely to be rated low on the Likert scale (as very similar), whereas when lines are stacked more towards the bottom, the cumulative probability is larger for higher values on the Likert scale (higher originality between structures compared).

When participants collaborated or built together but at different parts of the same structure, the final structures were more likely to be rated as very different (Figure 54).

Again, caution is required as the model prediction of the ratings at higher occasions of collaboration is very uncertain.

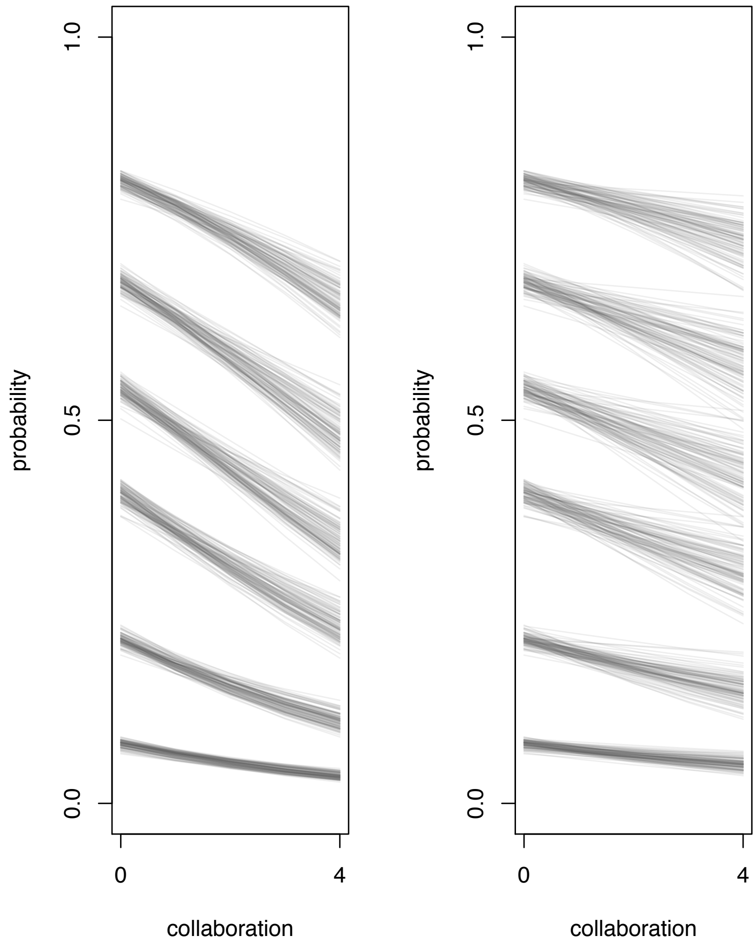


Figure 54: The posterior predictions of the ordered categorical models of participants building in pairs or groups in which there were a different number of events of collaboration (range 0 – 4). Each plot shows how the posterior probability distribution of predicted responses (ratings on the scale from 1 - very similar to 7 – very different) varies when more collaboration occurred in the pair or group members at the Interactive Research Pod in the cooperative condition. Since the raters rated two images at the same time, the left and right graph correspond to the pair being compared and the amount of collaboration that occurred when children built either the left or right image in the pair. For example, one image in the pair is set up to 0 counts of collaboration while the other image in the can have a range between 0 to 4. Lines show where predicted thresholds between the ordinal categories lie in terms of cumulative probability. Distributions show cumulative probability, when they are stacked more towards the top of the graph, this shows the structures compared were more likely to be rated low on the Likert scale (as very similar), whereas when lines are stacked more towards the bottom, the cumulative probability is larger for higher values on the Likert scale (higher originality between structures compared).

The results for the ratings that structures in pairs or groups that were discussing a lot had received were very uncertain. There is no clear change in the probability of the structures receiving ratings that would either suggest the structures are more likely to be rated as very different or very similar (Figure 55).

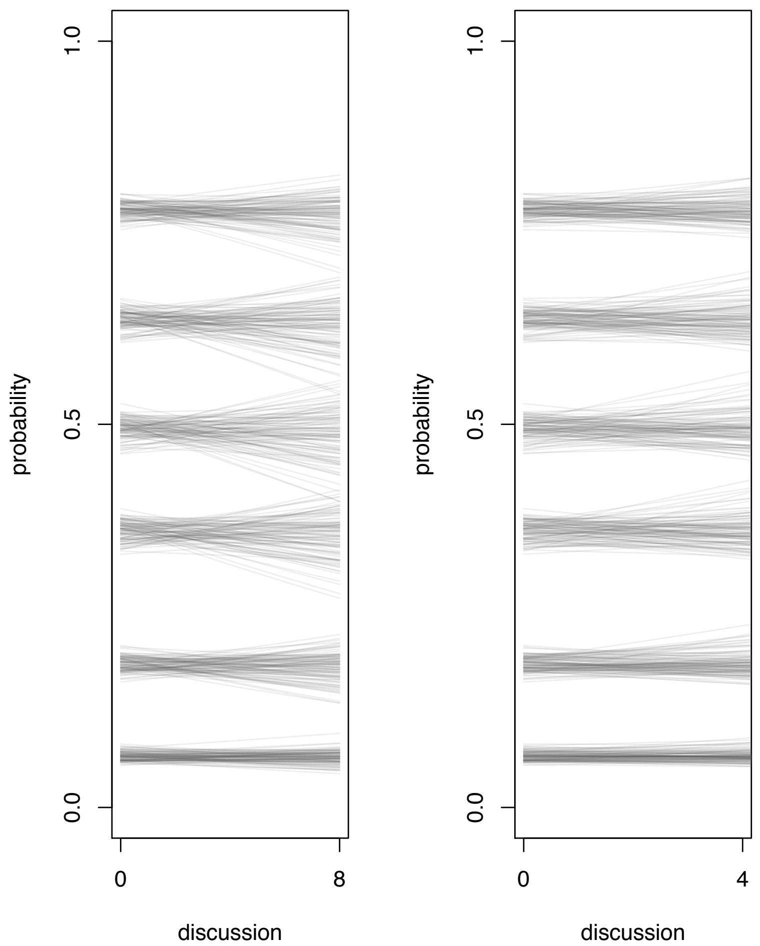


Figure 55: The posterior predictions of the ordered categorical models of participants building in pairs or groups in which there were different number of discussion events (range 0 – 8). Each plot shows how the posterior probability distribution of predicted responses (ratings on the scale from 1 - very similar to 7 – very different) varies when more discussions occurred in the pair or group members at the Interactive Research Pod in the cooperative condition. Since the raters rated two images at the same time, the left and right graph correspond to the pair being compared and the number of discussion events that occurred when children built either the left or right image in the pair. For example, one image in the pair is set up to 0 counts of discussion while the other image in the can have a range between 0 to 8. Lines show where predicted thresholds between the ordinal categories lie in terms of cumulative probability. Distributions show cumulative probability, when they are stacked more towards the top of the graph, this shows the structures compared were more likely to be rated low on the Likert scale (as very similar),

whereas when lines are stacked more towards the bottom, the cumulative probability is larger for higher values on the Likert scale (higher originality between structures compared).

The average age of pairs or groups did not clearly predict the originality of the structures built (Figure 56). However, the structures that older groups created were usually rated lower (lower originality) than the creations of groups that were on average younger, but the uncertainty is high.

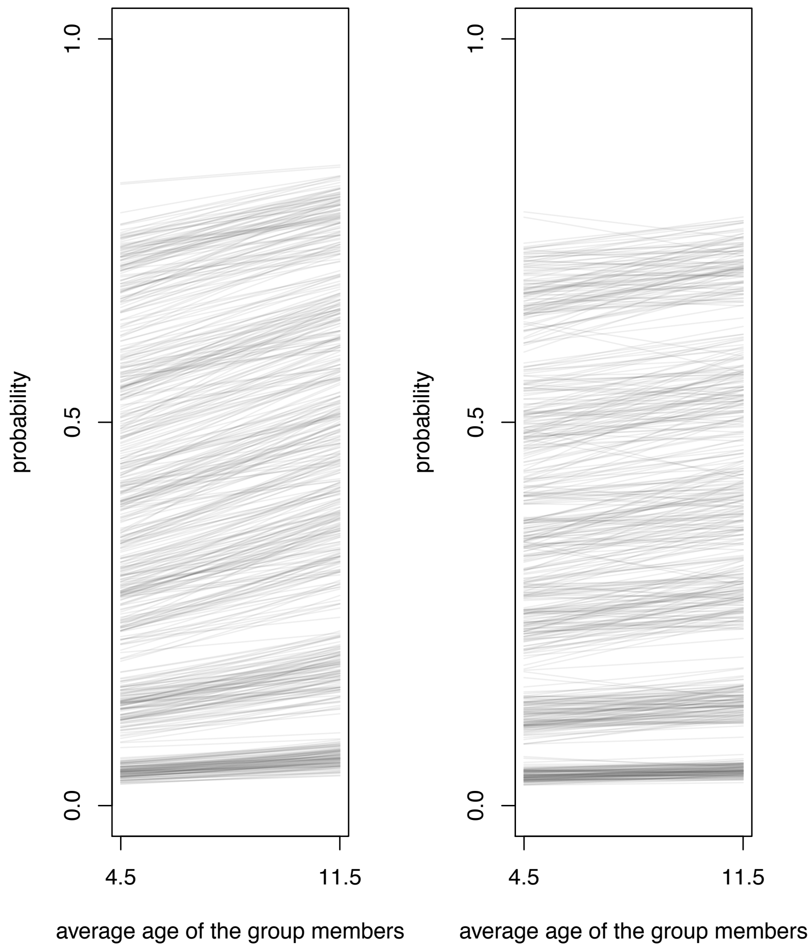


Figure 56: : The posterior predictions of the ordered categorical models of participants building in pairs or groups in which the average age (between 4.5 to 11.5 year old) of participants in the cooperative condition. Each plot shows how the posterior probability distribution of predicted responses (ratings on the scale from 1 - very similar to 7 – very different) varies when the average age of members in the

groups or pairs increases. Since the raters rated two images at the same time, the left and right graph correspond to the pair being compared and the average age of participants in the group at either the left or right image in the pair. For example, one image in the pair is set up to an average age of all groups while the other image can have a range between 4.5 to 11.5. Lines show where predicted thresholds between the ordinal categories lie in terms of cumulative probability. Distributions show cumulative probability, when they are stacked more towards the top of the graph, this shows the structures compared were more likely to be rated low on the Likert scale (as very similar), whereas when lines are stacked more towards the bottom, the cumulative probability is larger for higher values on the Likert scale (higher originality between structures compared).

5.4 Discussion

In this study I asked children to build with blocks in groups and I observed their behaviours, which enabled me to investigate copying, providing instructions, discussions, cooperation and collaboration in groups. By asking pairs and groups of children to build whatever they wanted from wooden blocks I could explore how any of the behaviours of interest impacted on the originality of their final structures.

The extent of instructing, adjusting and copying by an individual was not strongly influenced by age, sex and the number of friends individuals in groups had. These findings do not support my predictions, where I expected older children to instruct more and younger to copy more. The effect of age might not have been recorded, since the children in the study had a large range of ages.

In line with my prediction, the number of instructions increased with the average age of children in the group. This is probably due to the older children possessing better developed spatial language than younger children (Ferrara et al. 2011). However, I neglected other types of instructions children provided each other, such as physical demonstrations, which might be present especially in pairs and groups with younger participants (Ashley and Tomasello 2001). Age did not impact on copying behaviours in the groups. The uncertainty of the models predicting the amount of copying was high, therefore the results must be interpreted with caution. However, there was less copying in groups where the average age of children was low. This is probably due to the fact, that children did not have an older peer model who they could copy or even receive instructions from (Wood, Kendal, and Flynn 2015; Wood, Kendal, and Flynn 2012; Ashley and Tomasello 2001).

The amount of copying or discussing in pairs or groups did not impact on the overall creativity, or the originality rating of the structures children in pairs and groups built. This appear to correspond with my predictions and with studies that show children who copy do not innovate in extracting rewards from puzzle boxes (Carr, Kendal, and Flynn 2015; Wood, Kendal, and Flynn 2015). However, this study introduced an open-ended task and was not focused on problem solving, and thus innovation, but only on the overall originality of the structures children built. It appears that working on the same model resulted in diverse structures that were original.

Cooperation and collaboration within pairs or groups led to more original structures. This result corresponds to the finding that children exposed to a program that fostered learning through cooperation scored higher when individually tested on classical creativity tests than children who followed a program that was not specifically designed to facilitate cooperation (Baloche 1994). When children are cooperating, they provide others with space to develop and share their ideas and together create a final product. Typically, when cooperating there is no clear leader but instead the final product is the result of numerous ideas provided by different individuals. Since groups that cooperated and collaborated more built more original structures, it would be interesting to study its connection to social learning and creativity in more detail.

As I predicted, instructions limited the originality of the structures children built. It seems that when there were many instructions children did not explore the blocks as much as the groups with few instructions nor search for other structures to build. This corresponds to studies that found instructions limited exploration (Bonawitz et al. 2011). However, this would have to be experimentally tested by including a measurement of exploration when building with blocks and a control with individuals providing instructions being more or less proficient in completing the task. In the case of instructions, it would also be beneficial to follow whether the individuals in the pair or group that occupied the role of the leader restricted others from contributing their ideas. However, in the future the identification of the instructor might lead us to explore whether children express clear biases towards specific models. For instance, young children may perceive older children as being more knowledgeable and proficient in accomplishing the task and thus, more likely to follow to their instructions. Children use specific social learning strategies and follow the majority (Corriveau and Harris 2010, 2010; Morgan, Laland, and Harris 2015), older models in contexts such as formal learning environments (Rakoczy et al. 2010; Wood, Kendal, and Flynn 2012) and peers in playful environments (Ryalls, Gul, and Ryalls 2000; Nielsen, Cucchiaro, and Mohamedally 2012) and children prefer to copy innovations of peer demonstrators with assumed proficiency (Wood, Kendal, and Flynn 2015).

In contrast to my predictions and to previous studies (Beck et al. 2011; Sheridan et al. 2016; Neldner, Mushin, and Nielsen 2017; Beck et al. 2016; Cutting et al. 2014) that older children would create more original structures than younger children, pairs and groups that were on average older did not build less or more original structures than those that consisted of a younger average ages. In other studies innovation was tested through the use of problem-solving tasks, whereas in this study children were not limited by one solution. This may have enabled younger children to be perceived as equally creative as older children since the requirement for special knowledge was removed. Since toddlers also think divergently and are creative in the context of making jokes (Hoicka and Butcher 2016; Hoicka et al. 2016) the building block task used in this study may have been another natural way in which children express their creativity and thus, well suited for studying creativity in children of all age groups.

The findings are tentative since towards the higher counts of behaviours the results were uncertain. In this study there were not many groups in which the behaviour counts were high. This might be due to the fact these behaviours are actually rare or children do not verbally communicate much when building with blocks in free play. The latter has been shown in the past where parents who guided children through building with blocks used more spatial language than parents with children in free play (Borriello and Liben 2018; Ferrara et al. 2011).

Overall, the study exposed the connection between five types of behaviours children expressed when building in pairs or groups at the IRP with factors such as age and sex, and creativity measured through the comparison of pairs of structures children built together in pairs and groups. In particular, cooperating and collaborating when building together in pairs and groups seem to facilitate creativity in children. Therefore, experiments in naturalistic environments should complement studies of open diffusion in children with an implementation of tasks that can only be solved through cooperation and collaboration in groups (Ashley and Tomasello 2001; Warneken et al. 2011). All studied behaviours constitute cultural learning (Tomasello, Kruger, and Ratner 1993) children are exposed to in everyday life. Through teaching and everyday social experiences cooperation is maintained in human groups (Olson and Spelke 2008). I found social learning behaviours diversely impacted creativity in children but to understand their effects in further detail more controlled experiments with high experimental validity need to be conducted.

6 General Discussion

In this thesis I introduced a novel way of measuring creativity in an informal educational environment. I combined basic studies of social learning in children with innovation studies, but instead of using a goal oriented problem solving task I used open-ended tasks. In comparison to other experimental approaches and studies exploring children's innovation, the research presented in this thesis quantitatively explored children in naturalistic settings as well as within a quasi-experimental design. Typically, neither children nor adults are isolated from their social context (Legare, Sobel, and Callanan 2017; Legare, Gose, and Guess 2015; Flynn and Whiten 2010). Therefore, in order to gain a deeper understanding of cognition and learning, we need to combine various approaches and measure behaviours "in the wild" (Flynn and Whiten 2010), as is commonly reported in studies of non-human animal (Kendal, Galef, and Schaik 2010; Hill 2010).

With three different experimental designs in the science centre, I investigated cultural learning and its impact on creative behaviours in children in an open-ended task. First, I asked children to build spontaneously from wooden blocks. I was predominantly interested in exploring the effects instructions had on children's creativity. In this setting, I either provided children with direct instructions (through demonstrating affordances of the blocks), guided them through discovering the block's affordances with open questions, or did not provide any instructions. Overall, children's creativity did not differ among conditions. This did not support my prediction that instructions would hinder children's creativity. The sex and age of individuals did not impact on children's behaviours or the originality of their final structures.

Second, I collected data automatically, using the IRP, without actively being present at the research site. This enabled me to collect data in three conditions: social (transparent partitions) condition, asocial (opaque partitions but with the opportunity to 'peek' at others while building) condition and an asocial control (opaque partitions, different day) condition. This study design provided me an opportunity to explore naturalistic behaviours of children and explore how social learning affected their creativity. The differences among conditions were not large, the most original structures were in the asocial control condition where children were not able to see others building at the IRP or any previously built structures. This was in line with my predictions, though I did not foresee that even with opaque partitions set up in the asocial condition, children would still build similar structures to others building at the IRP. The sex and age of the participants did not impact on the originality of their final builds.

Third, I investigated how instructions, copying and cooperation affected the final structures children in pairs and groups built. A research assistant and I recruited pairs and groups of children to build with blocks. Initially, I focused on behaviours of individual children, such as instructing, copying and adjusting what others added to the structure. These behaviours could not be explained by factors such as group composition in relation to the sex and the age of individuals. Then, I focused on group behaviours such as the number of instructions provided by individuals in the group, the number of times they worked together on the same parts of the structure (cooperation), the number of times they worked on different parts of the same structure (collaboration), the number of times individuals in groups discussed their building and the number of times any individual in the group copied another individual. In groups that were on average older more instructions were present. When children cooperated and collaborated in building the same structure, the final creations were more original than when they were instructing more or when they just copied or discussed the structures. Thus, cooperation and collaboration would appear to be important components of creativity, which are commonly neglected in studies of innovation of children.

6.1 Creativity and cumulative culture in an open-ended task

Overall, following the above described research design, I expected children's creativity to be lower in conditions when they were either instructed or could imitate others and that children would build more original structures when they created on their own. Throughout the thesis it became clear that this was not true and being guided or being exposed to other models did not necessarily limit children's creativity in an open-ended task. This further supports the idea that lone geniuses and creatives do not necessarily exist, but rather, new knowledge and innovations emerge from collective knowledge (Muthukrishna and Henrich 2016).

Arguably, open-endedness is an integral part of cumulative culture, since the ratchet effect supposedly leads to technology, ideas or institutions, that no individual could come up with in their lifetime (Tennie, Caldwell, and Dean 2018 as cited in Reindl and Tennie 2018). We can understand the path to innovations and solutions to problems as open-ended and the pool of possible solutions is large (for example there is no one way of solving climate change) (Moreau and Engeset 2015). However, focusing on innovations in retrospect and from the point of view of historic creativity, it might seem there was - and still is - only one 'correct' solution to a problem. Therefore, studies of creativity, innovation and problem-solving also need to incorporate open-ended tasks.

With the use of open-ended tasks a richer picture of how the complex and more efficient solutions that we see in human culture will emerge. In this thesis I have found connections between three types of cultural learning: direct instructions, imitation and cooperation, and children's creativity. However, none of which dampened creativity in children. As children grow they need to learn how to combine all learning strategies with becoming sufficient in solving everyday problems and tasks (Gopnik et al. 2017; Legare and Harris 2016).

Studies presented in this thesis included a wide age range of children and age was only used as a controlling factor. In most cases age did not correlate with children's creativity. This was expected since numerous studies have shown children express different types of exploration (Bonawitz et al. 2011), divergent thinking (Hoicka and Bijvoet-van den Berg 2013), making up jokes (Hoicka and Akhtar 2011) and innovation of chimpanzee tools (Reindl et al. 2016) at ages younger than 4 years. Age only correlated with children's behaviours of block manipulations (Chapter 3) and instances of copying and instructions in groups of children (Chapter 5). However, the strength of these relationships was weak and not as strong as in studies performed in controlled environments. Which might be due to the fact I used an open-ended task, without clear goal. Further investigation of children's behaviours and learning when using an open-ended task could be performed in controlled settings.

6.2 Science centre

Science centres offer a great place to explore children's cognition and learning through a multitude of ways (Jipson and Sobel 2015; Callanan 2012). First, the children from non-WEIRD cultures are usually neglected in social learning and innovation studies with a few exceptions (Nielsen and Tomaselli 2010; Nielsen et al. 2016; Clegg and Legare 2016b; Henrich, Heine, and Norenzayan 2010). Science centres offer a pool of participants from diverse backgrounds (Callanan, Martin, and Luce 2015), albeit the LSC has an entrance fee which might limit the visitors pool. Second, children are prone to support and show enthusiasm to cooperate when participating in studies in science centres (Corriveau et al. 2015; Legare, Gose, and Guess 2015). Participation might also foster children's interest in science and present them with different ways of conducting research that might not have previously been presented in the science centres (for example social learning in children) (Corriveau et al. 2015). Third, researchers have the opportunity to present their studies to the general public and enhance their public engagement skills (Callanan, Martin, and Luce 2015; Corriveau et al. 2015). Fourth, studies presented in this thesis can inform science centres about the use of instructions and the position of exhibits depending on their aim. For example when not aiming to foster creativity, exhibits could be positioned in view of each other but when supporting development of creativity they should be more secluded to encourage discovery (asocial) learning rather than social learning. Furthermore, they could use research-based approaches to explore how best to inform and educate different audiences about science.

Some practical constraints of conducting research within the Life Science Centre have emerged when conducting experiments presented in this thesis. First, it was not possible to recruit school children since guardian consent was necessary and with time being limited, I could not set up contact with schools to propose cooperation. Second, LSC has very sporadic visits. During school holidays LSC is very busy making it challenging to establish conditions which are good representations of quasi-naturalistic experiments and at other times there are not enough visitors that would want to participate in the studies. Third, it is challenging to keep the children's focus only on the task at hand since other exhibits and activities in the science centre can be distracting. This is especially true in the studies presented in Chapters 4 and 5 where the experimenter was not present. Likewise, when an experimenter is not present it is not possible to make sure that the required number of children built at the same time (variability in this fact had to be controlled for in analyses instead). The same is true for studies presented in Chapter 5 where two experimenters were present but there were no partitions separating the Interactive Research Pod from the rest of the exhibits.

6.3 The Interactive Research Pod

The IRP has proved to be a suitable exhibit to study social learning. Visitors were attracted to it which could be attributed to social facilitation when individuals show an interest in exhibits that others are engaged with (Zajonc 1965). The data sets I received through the use of the IRP were large, but also messy. Therefore, advanced modelling is necessary to explain behaviours recorded at the IRP. In the future the exhibit could be used to study social learning strategies (Rendell et al. 2011; Kendal et al. 2018; Laland 2004), transmission of knowledge between generations (Haden et al. 2015; Demps et al. 2012), ratchet effect (Tomasello, Kruger, and Ratner 1993; Tomasello 1999; Tennie, Call, and Tomasello 2009; Tennie et al. 2014; Eva Reindl and Tennie 2018) and spatial skills in children (Casey and Bobb 2003; Casey et al. 2008). The exhibit can easily be modified to specific research questions and a more controlled way of measuring creativity and innovation. Researchers could explore how visitors interact and tackle specific problems, that can be solved through innovation or social learning, in a similar way to a study performed in the past in a Zoo (Macdonald and Whiten 2011; Whiten et al. 2016).

6.4 Bayesian methods

Naturalistic observations and quasi experiments lead to data that can be messy (Flynn and Whiten 2010), but novel analytical techniques (statistical modelling and automatic data collection) can make it easier to collect and analyse. Science has been biased towards positive results (Smaldino and McElreath 2016), mainly relying on the p-value showing the significant differences between samples (Open Science Collaboration 2015; Colquhoun 2017; Wasserstein and Lazar 2016). Here I used Bayesian statistics and presented a step away from the dichotomy of positive versus negative results but focused instead on the strength of the findings. Data which include a lot of parameters that need to be accounted for are better explored through multiple competing predictions (Frankenhuis and Nettle 2018). Bayesian analysis offers model selection and model averaging for explaining phenomenon of interest (Symonds and Moussalli 2011).

In this study social learning sometimes hindered children's creativity, but the effect was mild. The analysis I used offered me a way to explore a diverse range of behaviours as well as accounting for possible differences between sexes and ages of children. I presented results from the models that included multiple parameters of interest. Therefore, it enabled me to see a complex image of the intersection between different types of social learning and creativity. Mostly the results in this thesis were uncertain and did not follow the predictions. However, they are still valuable for future studies in science centres and might provide more informative priors for further Bayesian analysis (Etz and Vandekerckhove 2018; Etz et al. 2018).

6.5 Limitations of the study design

In all studies presented in this thesis, children of a wide age range participated. Unfortunately, as a result of the wide age range I could not say anything about specific changes in developmental stages of children's development. However, I was able to explore whether the behaviours expressed were in line with predictions originating from the studies performed in more controlled environment.

I did not know the motivations of the children participating in the studies and whether they understood the aim of the task or if they were even trying to be creative. Motivation to participate in research, learn from others and be creative is highly neglected and should be considered in the future as it has been in studies inquiry-based learning in children (Van Schijndel, Jansen, and Raijmakers 2018). If the environment was fairly novel they might have been preoccupied with exploring the place and were not highly motivated to build with blocks for the purpose of research, especially when researchers were not present. Although, children are usually encouraged by parents to participate and receive an award for participation when researchers are present which helps the child remain focused on the task (Corriveau et al. 2015).

The age ranges of children included in the studies were large and therefore the effect of age might exist but the data for each age category was sparse which might have resulted in a weak affect. Previous experience at and familiarity with the science centre surely affected what children built and how they approached the tasks. It is not possible to generalise the effect of social learning on creativity in children since the visitors to the

LSC mainly come from the North East of the United Kingdom and the diversity is not as great as in some science centres around the globe, especially as the entrance fee is pricey. I also recognise that data regarding children's ethnicity and their social economic status may have provided a better and more detailed view on the data used in this study and make studies presented in this thesis more reproducible. Therefore, studies using similar approaches should be performed to see whether results replicate in different contexts.

6.6 Future directions

In the future more controlled experiments of social learning and creativity in children could be performed with the help of school groups where researchers would have basic information about the members of the groups. When recruiting children who visit the LSC in school groups we also gain more diverse samples and do not rely on parents taking their children and covering the entrance fee.

With the opportunity to have more similarly aged participants we could explore the changes in behaviours among specific ages. However, with the samples of participants in these studies I could not do this because of the wide age ranges of participants. Furthermore, controlled replication of the findings should be carried out in a closed room where only an experimenter would be present with an individual in study one (Chapter 3), three individuals in study 2 (Chapter 4) and pairs and groups of individuals in study 3 (Chapter 5). This way there would not be any distracting stimuli from other people and exhibits and we could identify any factors we might have overlooked when performing studies when others were present at and around the exhibits. Additionally, the IRP was developed as a science centre exhibit and could be produced in the same way in another science centre around the globe. Simultaneous collection of the data in two or more places would make it easier to test whether findings replicate in other science centres. Ideally, we could also develop a transferable version of the IRP with which we could also test children's creativity and social learning outside WEIRD populations.

Studies in naturalistic settings should blossom nowadays where better analytical and data collection methods provide us with an opportunity to collect and analyse large data sets. These might lead to a better understanding of children's behaviour as well as give museum practitioners new data which are simply not obtainable through the use of surveys and focus groups (Falk 2009; Andre, Durksen, and Volman 2017).

Furthermore, we could explore social learning biases in naturalistic settings by selecting different data from the Interactive research pod. For example, we could collect data where we can record copying between generations (e.g. through videos where adults and children provide consent or assent, respectively, to use their data) or focus on social learning in adults or a specific age groups to which we have the data.

6.7 Final remarks

If we want to fully understand social learning strategies employed in different contexts, as well as their influence on innovation in children, we need to carry out more studies in naturalistic environments that children use and occupy daily. We need to develop tasks that children are faced with in daily life and in which they can learn from others, but nevertheless also develop and express original behaviours.

The present study was an attempt to combine studies of innovation and social learning in children together with their activities in an informal learning environment. Studies that are more ecologically valid can help to inform controlled studies with high experimental validity about what experimental designs are most appropriate for exploring how children behave. Furthermore, observing how children naturally innovate and create might enable researchers to design methods that move away from the standard battery of tests that are currently used to study creativity and innovation in children. Thus, making researchers themselves more innovative and creative.

7 Bibliography

Abraham, Anna. 2019. "What Is Creativity?" In *The Neuroscience of Creativity*, 1–10. Cambridge Fundamentals of Neuroscience in Psychology. Cambridge, United Kingdom ; New York, NY: Cambridge University Press.

Andre, Lucija, Tracy Durksen, and Monique L. Volman. 2017. "Museums as Avenues of Learning for Children: A Decade of Research." *Learning Environments Research* 20 (1): 47–76. https://doi.org/10.1007/s10984-016-9222-9.

Ashley, Jennifer, and Michael Tomasello. 2001. "Cooperative Problem-Solving and Teaching in Preschoolers." *Social Development* 7 (2): 143–63. https://doi.org/10.1111/1467-9507.00059.

Bailey-Ross, C., H. Rudman, J. Kendal, Z. Mursic, A. Lloyd, B. Ross, and R. L. Kendal. 2018. "Reconnecting Epistemologies via Co-Design and Participatory Action Research Practice." In *Beyond Disciplinarity in Social Research : Methodologies, Epistemologies and Philosophies.*, edited by C. Hayes, J. Fulton, and K. Petrie. London: Routledge. http://dro.dur.ac.uk/23927/.

Baloche, Lynda. 1994. "Creativity and Cooperation in the Elementary Music Classroom*." *The Journal of Creative Behavior* 28 (4): 255–65. https://doi.org/10.1002/j.2162-6057.1994.tb00732.x.

Bateson, Patrick, and Paul Martin. 2013. *Play, Playfulness, Creativity and Innovation*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139057691.

Beck, Sarah R., Ian A. Apperly, Jackie Chappell, Carlie Guthrie, and Nicola Cutting. 2011. "Making Tools Isn't Child's Play." *Cognition* 119 (2): 301–6. https://doi.org/10.1016/j.cognition.2011.01.003.

Beck, Sarah R., Clare Williams, Nicola Cutting, Ian A. Apperly, and Jackie Chappell. 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: Biological Sciences* 371 (1690): 20150190. https://doi.org/10.1098/rstb.2015.0190.

Beghetto, Ronald A., and James C. Kaufman. 2007. "Toward a Broader Conception of Creativity: A Case for 'Mini-c' Creativity." *Psychology of Aesthetics, Creativity, and the Arts* 1 (2): 73–79. https://doi.org/10.1037/1931-3896.1.2.73.

Bijvoet-van den Berg, Simone, and Elena Hoicka. 2014. "Individual Differences and Age-Related Changes in Divergent Thinking in Toddlers and Preschoolers." *Developmental Psychology* 50 (6): 1629–39. https://doi.org/10.1037/a0036131.

Bird, C. D., and N. J. Emery. 2009. "Insightful Problem Solving and Creative Tool Modification by Captive Nontool-Using Rooks." *Proceedings of the National Academy of Sciences* 106 (25): 10370–75. https://doi.org/10.1073/pnas.0901008106.

Boden, Margaret A. 2004. *The Creative Mind: Myths and Mechanisms*. 2nd ed. London: Routledge.

Bonawitz, Elizabeth Baraff, Tessa J.P. van Schijndel, Daniel Friel, and Laura Schulz. 2012. “Children Balance Theories and Evidence in Exploration, Explanation, and Learning.” *Cognitive Psychology* 64 (4): 215–34. https://doi.org/10.1016/j.cogpsych.2011.12.002.

Bonawitz, Elizabeth, Patrick Shafto, Hyowon Gweon, Noah D. Goodman, Elizabeth Spelke, and Laura Schulz. 2011. “The Double-Edged Sword of Pedagogy: Instruction Limits Spontaneous Exploration and Discovery.” *Cognition* 120 (3): 322–30. https://doi.org/10.1016/j.cognition.2010.10.001.

Borriello, Giulia A., and Lynn S. Liben. 2018. “Encouraging Maternal Guidance of Preschoolers’ Spatial Thinking During Block Play.” *Child Development* 89 (4): 1209–22. https://doi.org/10.1111/cdev.12779.

Boyd, Robert, Peter J. Richerson, and Joseph Henrich. 2011. “The Cultural Niche: Why Social Learning Is Essential for Human Adaptation.” *Proceedings of the National Academy of Sciences* 108 (Supplement 2): 10918–25. https://doi.org/10.1073/pnas.1100290108.

Boyette, Adam H. 2016. “Children’s Play and the Integration of Social and Individual Learning: A Cultural Niche Construction Perspective.” In *Social Learning and Innovation in Contemporary Hunter-Gatherers*, edited by H. Terrashima and Barry S. Hewlett, 159–69. Springer Japan.

Buchsbaum, Daphna, Alison Gopnik, Thomas L. Griffiths, and Patrick Shafto. 2011. “Children’s Imitation of Causal Action Sequences Is Influenced by Statistical and Pedagogical Evidence.” *Cognition* 120 (3): 331–40. https://doi.org/10.1016/j.cognition.2010.12.001.

Burdett, Emily R. R., Amanda J. Lucas, Daphna Buchsbaum, Nicola McGuigan, Lara A. Wood, and Andrew Whiten. 2016. “Do Children Copy an Expert or a Majority? Examining Selective Learning in Instrumental and Normative Contexts.” *PLOS ONE* 11 (10): e0164698. https://doi.org/10.1371/journal.pone.0164698.

Bürkner, Paul, and Matti Vuorre. 2018. “Ordinal Regression Models in Psychological Research: A Tutorial.” *PsyArXiv Preprints*. https://doi.org/10.31234/osf.io/x8swp.

Bürkner, Paul-Christian. 2017a. “Brms: An R Package for Bayesian Multilevel Models Using Stan.” *Journal of Statistical Software* 80 (1). https://doi.org/10.18637/jss.v080.i01.

———. 2017b. “Advanced Bayesian Multilevel Modeling with the R Package Brms.” *ArXiv:1705.11123 [Stat]*, May. http://arxiv.org/abs/1705.11123.

Butler, Lucas P., and Ellen M. Markman. 2012. “Preschoolers Use Intentional and Pedagogical Cues to Guide Inductive Inferences and Exploration: Pedagogical Cues and Exploration.” *Child Development* 83 (4): 1416–28. https://doi.org/10.1111/j.1467-8624.2012.01775.x.

Caldwell, C. A, and A. E Millen. 2008. “Studying Cumulative Cultural Evolution in the Laboratory.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 363 (1509): 3529–39. https://doi.org/10.1098/rstb.2008.0133.

Caldwell, Christine A., and Ailsa E. Millen. 2008. "Experimental Models for Testing Hypotheses about Cumulative Cultural Evolution." *Evolution and Human Behavior* 29 (3): 165–71. https://doi.org/10.1016/j.evolhumbehav.2007.12.001.

———. 2010. "Conservatism in Laboratory Microsocieties: Unpredictable Payoffs Accentuate Group-Specific Traditions." *Evolution and Human Behavior* 31 (2): 123–30. https://doi.org/10.1016/j.evolhumbehav.2009.08.002.

Call, Josep, Malinda Carpenter, and Michael Tomasello. 2005. "Copying Results and Copying Actions in the Process of Social Learning: Chimpanzees (Pan Troglodytes) and Human Children (Homo Sapiens)." *Animal Cognition* 8 (3): 151–63. https://doi.org/10.1007/s10071-004-0237-8.

Callanan, Maureen A. 2012. "Conducting Cognitive Developmental Research in Museums: Theoretical Issues and Practical Considerations." *Journal of Cognition and Development* 13 (2): 137–51. https://doi.org/10.1080/15248372.2012.666730.

Callanan, Maureen, Jennifer Martin, and Megan Luce. 2015. "Two Decades of Families Learning in a Children's Museum: A Partnership of Research and Exhibit Development." In *Cognitive Development in Museum Settings: Relating Research and Practice*, 1st ed., 15–35. Routledge.

Carpenter, Malinda, Nameera Akhtar, and Michael Tomasello. 1998. "Fourteen- through 18-Month-Old Infants Differentially Imitate Intentional and Accidental Actions." *Infant Behavior and Development* 21 (2): 315–30. https://doi.org/10.1016/S0163-6383(98)90009-1.

Carpenter, Malinda, Josep Call, and Michael Tomasello. 2005. "Twelve- and 18-Month-Olds Copy Actions in Terms of Goals." *Developmental Science* 8 (1): F13–20. https://doi.org/10.1111/j.1467-7687.2004.00385.x.

Carr, Kayleigh, Rachel L. Kendal, and Emma G. Flynn. 2015. "Imitate or Innovate? Children's Innovation Is Influenced by the Efficacy of Observed Behaviour." *Cognition* 142 (September): 322–32. https://doi.org/10.1016/j.cognition.2015.05.005.

———. 2016. "Eureka!: What Is Innovation, How Does It Develop, and Who Does It?" *Child Development* 87 (5): 1505–19. https://doi.org/10.1111/cdev.12549.

Carruthers, Peter. 2002. "Human Creativity: Its Cognitive Basis, Its Evolution, and Its Connections with Childhood Pretence." *The British Journal for the Philosophy of Science* 53 (2). http://www.jstor.org/stable/3541765.

Casey, Beth, and Barbara Bobb. 2003. "The Power of Block Building." *Teaching Children Mathematics* 10 (2): 98–102. https://www.jstor.org/stable/41198090.

Casey, Beth M., Nicole Andrews, Holly Schindler, Joanne E. Kersh, Alexandra Samper, and Juanita Copley. 2008. "The Development of Spatial Skills Through Interventions Involving Block Building Activities." *Cognition and Instruction* 26 (3): 269–309. https://doi.org/10.1080/07370000802177177.

Clay, Zanna, Harriet Over, and Claudio Tennie. 2018. "What Drives Young Children to Over-Imitate? Investigating the Effects of Age, Context, Action Type, and Transitivity." *Journal of Experimental Child Psychology* 166 (February): 520–34. https://doi.org/10.1016/j.jecp.2017.09.008.

Clay, Zanna, and Claudio Tennie. 2017. "Is Overimitation a Uniquely Human Phenomenon? Insights From Human Children as Compared to Bonobos." *Child Development*, July. https://doi.org/10.1111/cdev.12857.

Clegg, Jennifer M., and Cristine H. Legare. 2016a. "Instrumental and Conventional Interpretations of Behavior Are Associated With Distinct Outcomes in Early Childhood." *Child Development* 87 (2): 527–42. https://doi.org/10.1111/cdev.12472.

———. 2016b. "A Cross-Cultural Comparison of Children's Imitative Flexibility." *Developmental Psychology* 52 (9): 1435–44. https://doi.org/10.1037/dev0000131.

Collaboration, Open Science. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349 (6251): aac4716. https://doi.org/10.1126/science.aac4716.

Colquhoun, David. 2017. "The Reproducibility Of Research And The Misinterpretation Of P Values," October. https://doi.org/10.1101/144337.

Corriveau, Kathleen H., and Paul L. Harris. 2010. "Preschoolers (Sometimes) Defer to the Majority in Making Simple Perceptual Judgments." *Developmental Psychology* 46 (2): 437–45. https://doi.org/10.1037/a0017553.

Corriveau, Kathleen H., Rebecca Kipling, Samuel Ronfard, Marta Biarnes, Brittany M. Jeye, and Paul L. Harris. 2015. "The Living Laboratory Model: A Mutual Professional Development Model for Museum-Based Research Partneship." In *Cognitive Development in Museum Settings: Relating Research and Practice*, 1st ed., 65–83. Routledge.

Crowley, Kevin, Maureen A. Callanan, Jennifer L. Jipson, Jodi Galco, Karen Topping, and Jeff Shrager. 2001. "Shared Scientific Thinking in Everyday Parent-Child Activity." *Science Education* 85 (6): 712–32. https://doi.org/10.1002/sce.1035.

Csibra, Gergely, and György Gergely. 2009. "Natural Pedagogy." *Trends in Cognitive Sciences* 13 (4): 148–53. https://doi.org/10.1016/j.tics.2009.01.005.

———. 2011. "Natural Pedagogy as Evolutionary Adaptation." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 366 (1567): 1149–57. https://doi.org/10.1098/rstb.2010.0319.

Cutting, Nicola, Ian A. Apperly, and Sarah R. Beck. 2011. "Why Do Children Lack the Flexibility to Innovate Tools?" *Journal of Experimental Child Psychology* 109 (4): 497–511. https://doi.org/10.1016/j.jecp.2011.02.012.

Cutting, Nicola, Ian A. Apperly, Jackie Chappell, and Sarah R. Beck. 2014. "The Puzzling Difficulty of Tool Innovation: Why Can't Children Piece Their Knowledge Together?" *Journal of Experimental Child Psychology* 125 (September): 110–17. https://doi.org/10.1016/j.jecp.2013.11.010.

Dean Jr., David, and Deanna Kuhn. 2007. "Direct Instruction vs. Discovery: The Long View." *Science Education* 91 (3): 384–97. https://doi.org/10.1002/sce.20194.

Dean, L. G., R. L. Kendal, S. J. Schapiro, B. Thierry, and K. N. Laland. 2012. "Identification of the Social and Cognitive Processes Underlying Human Cumulative Culture." *Science* 335 (6072): 1114–18. https://doi.org/10.1126/science.1213969.

Dean, Lewis G., Gill L. Vale, Kevin N. Laland, Emma Flynn, and Rachel L. Kendal. 2014. "Human Cumulative Culture: A Comparative Perspective." *Biological Reviews* 89 (2): 284–301. https://doi.org/10.1111/brv.12053.

Demps, Kathryn, Francisco Zorondo-Rodríguez, Claude García, and Victoria Reyes-García. 2012. "Social Learning across the Life Cycle: Cultural Knowledge Acquisition for Honey Collection among the Jenu Kuruba, India." *Evolution and Human Behavior* 33 (5): 460–70. https://doi.org/10.1016/j.evolhumbehav.2011.12.008.

Derex, Maxime, and Robert Boyd. 2015. "The Foundations of the Human Cultural Niche." *Nature Communications* 6 (September): 8398. https://doi.org/10.1038/ncomms9398.

Diedrich, Jennifer, Mathias Benedek, Emanuel Jauk, and Aljoscha C. Neubauer. 2015. "Are Creative Ideas Novel and Useful?" *Psychology of Aesthetics, Creativity, and the Arts* 9 (1): 35–40. https://doi.org/10.1037/a0038688.

DiYanni, Cara, Deniela Nini, and Whitney Rheel. 2011. "Looking Good versus Doing Good: Which Factors Take Precedence When Children Learn about New Tools?" *Journal of Experimental Child Psychology* 110 (4): 575–91. https://doi.org/10.1016/j.jecp.2011.06.002.

Etz, Alexander, Quentin F. Gronau, Fabian Dablander, Peter A. Edelsbrunner, and Beth Baribault. 2018. "How to Become a Bayesian in Eight Easy Steps: An Annotated Reading List." *Psychonomic Bulletin & Review* 25 (1): 219–34. https://doi.org/10.3758/s13423-017-1317-5.

Etz, Alexander, and Joachim Vandekerckhove. 2016. "A Bayesian Perspective on the Reproducibility Project: Psychology." Edited by Daniele Marinazzo. *PLOS ONE* 11 (2): e0149794. https://doi.org/10.1371/journal.pone.0149794.

———. 2018. "Introduction to Bayesian Inference for Psychology." *Psychonomic Bulletin & Review* 25 (1): 5–34. https://doi.org/10.3758/s13423-017-1262-3.

Evans, Cara L., Kevin N. Laland, Malinda Carpenter, and Rachel L. Kendal. 2018. "Selective Copying of the Majority Suggests Children Are Broadly 'Optimal-' Rather than 'over-' Imitators." *Developmental Science* 0 (0): e12637. https://doi.org/10.1111/desc.12637.

Falk, John H. 2009. *Identity and the Museum Visitor Experience*. Left Coast Press.

Ferrara, Katrina, Kathy Hirsh-Pasek, Nora S. Newcombe, Roberta Michnick Golinkoff, and Wendy Shallcross Lam. 2011. "Block Talk: Spatial Language During Block Play." *Mind, Brain, and Education* 5 (3): 143–51. https://doi.org/10.1111/j.1751-228X.2011.01122.x.

Flynn, E. 2008. "Investigating Children as Cultural Magnets: Do Young Children Transmit Redundant Information along Diffusion Chains?" *Philosophical Transactions of the Royal Society B: Biological Sciences* 363 (1509): 3541–51. https://doi.org/10.1098/rstb.2008.0136.

Flynn, Emma G., Kevin N. Laland, Rachel L. Kendal, and Jeremy R. Kendal. 2013. "Target Article with Commentaries: Developmental Niche Construction." *Developmental Science* 16 (2): 296–313. https://doi.org/10.1111/desc.12030.

Flynn, Emma, Cameron Turner, and Luc-Alain Giraldeau. 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: Biological Sciences* 371 (1690): 20150189. https://doi.org/10.1098/rstb.2015.0189.

———. 2018. "Follow (or Don't Follow) the Crowd: Young Children's Conformity Is Influenced by Norm Domain and Age." *Journal of Experimental Child Psychology* 167 (March): 222–33. https://doi.org/10.1016/j.jecp.2017.10.014.

Flynn, Emma, and Andrew Whiten. 2008. "Cultural Transmission of Tool Use in Young Children: A Diffusion Chain Study." *Social Development* 17 (3): 699–718. https://doi.org/10.1111/j.1467-9507.2007.00453.x.

———. 2010. "Studying Children's Social Learning Experimentally 'in the Wild.'" *Learning & Behavior* 38 (3): 284–96. https://doi.org/10.3758/LB.38.3.284.

———. 2012. "Experimental 'Microcultures' in Young Children: Identifying Biographic, Cognitive, and Social Predictors of Information Transmission." *Child Development* 83 (3): 911–25. https://doi.org/10.1111/j.1467-8624.2012.01747.x.

Fogarty, L., P. Strimling, and K. N. Laland. 2011. "The Evolution of Teaching." *Evolution* 65 (10): 2760–70. https://doi.org/10.1111/j.1558-5646.2011.01370.x.

Fogarty, Laurel, Nicole Creanza, and Marcus W. Feldman. 2015. "Cultural Evolutionary Perspectives on Creativity and Human Innovation." *Trends in Ecology & Evolution* 30 (12): 736–54. https://doi.org/10.1016/j.tree.2015.10.004.

Frankenhuis, Willem E., and Daniel Nettle. 2018. "Open Science Is Liberating and Can Foster Creativity." *Perspectives on Psychological Science* 13 (4): 439–47. https://doi.org/10.1177/1745691618767878.

Frick, Aurélien, Fabrice Clément, and Thibaud Gruber. 2017. "Evidence for a Sex Effect during Overimitation: Boys Copy Irrelevant Modelled Actions More than Girls across Cultures." *Royal Society Open Science* 4 (12): 170367. https://doi.org/10.1098/rsos.170367.

Fullerton, Andrew S. 2009. "A Conceptual Framework for Ordered Logistic Regression Models." *Sociological Methods & Research* 38 (2): 306–47. https://doi.org/10.1177/0049124109346162.

Gamer, Matthias, Jim Lemon, and Ian Fellows Puspendra Singh <puspendra.pusp22@gmail.com>. 2019. *Irr: Various Coefficients of Interrater Reliability and Agreement*. https://CRAN.R-project.org/package=irr.

Garaigordobil, Maite. 2006. "Intervention in Creativity With Children Aged 10 and 11 Years: Impact of a Play Program on Verbal and Graphic–Figural Creativity." *Creativity Research Journal* 18 (3): 329–45. https://doi.org/10.1207/s15326934crj1803_8.

Gartenhaus, A. 1997. *Minds in Motion: Using Museums to Expand Creative Thinking*. San Francisco: Caddo Gap Press.

Gentner, Dedre, Susan C. Levine, Raedy Ping, Ashley Isaia, Sonica Dhillon, Claire Bradley, and Garrett Honke. 2016. "Rapid Learning in a Children's Museum via Analogical Comparison." *Cognitive Science* 40 (1): 224–40. https://doi.org/10.1111/cogs.12248.

Gergely, György, Harold Bekkering, and Ildikó Király. 2002. "Developmental Psychology: Rational Imitation in Preverbal Infants." *Nature* 415 (6873): 755. https://doi.org/10.1038/415755a.

German, Tim P., and Margaret Anne Defeyter. 2000. "Immunity to Functional Fixedness in Young Children." *Psychonomic Bulletin & Review* 7 (4): 707–12. https://doi.org/10.3758/BF03213010.

Göckeritz, Susanne, Marco F.H. Schmidt, and Michael Tomasello. 2014. "Young Children's Creation and Transmission of Social Norms." *Cognitive Development* 30 (April): 81–95. https://doi.org/10.1016/j.cogdev.2014.01.003.

Golemiec, Mireille, Jonathan Schneider, W. Thomas Boyce, Nicole R. Bush, Nancy Adler, and Joel D. Levine. 2016. "Layered Social Network Analysis Reveals Complex Relationships in Kindergarteners." *Frontiers in Psychology* 7 (March). https://doi.org/10.3389/fpsyg.2016.00276.

Gopnik, A. 2012. "Scientific Thinking in Young Children: Theoretical Advances, Empirical Research, and Policy Implications." *Science* 337 (6102): 1623–27. https://doi.org/10.1126/science.1223416.

Gopnik, Alison, Shaun O'Grady, Christopher G. Lucas, Thomas L. Griffiths, Adrienne Wente, Sophie Bridgers, Rosie Aboody, Hoki Fung, and Ronald E. Dahl. 2017. "Changes in Cognitive Flexibility and Hypothesis Search across Human Life History from Childhood to Adolescence to Adulthood." *Proceedings of the National Academy of Sciences* 114 (30): 7892–99. https://doi.org/10.1073/pnas.1700811114.

Gräfenhain, Maria, Tanya Behne, Malinda Carpenter, and Michael Tomasello. 2009. "Young Children's Understanding of Joint Commitments." *Developmental Psychology* 45 (5): 1430–43. https://doi.org/10.1037/a0016122.

Grusec, Joan E., and Rona Abramovitch. 1982. "Imitation of Peers and Adults in a Natural Setting: A Functional Analysis." *Child Development* 53 (3): 636. https://doi.org/10.2307/1129374.

Guilford, J. P. 1959. "Three Faces of Intellect." *American Psychologist* 14 (8): 469–79. https://doi.org/doi:10.1037/h0046827.

———. 1967. *The Nature of Human Intelligence*. New York, NY, US: McGraw-Hill.

Guilford, J. P., P. R. Christensen, P. R. Merrifield, and R. C. Wilson. 1978. *Alternate Uses: Manual of Instructions and Interpretations*. Orange, CA: Sheridan Psychological Services.

Haden, Catherine A., Tsivia Cohen, David H. Uttal, and Maria Marcus. 2015. "Building Learning: Narrating Experiences in a Children's Museum." In *Cognitive Development in Museum Settings: Relating Research and Practice*, 84–103. Routledge.

Hamann, Katharina, Felix Warneken, Julia R. Greenberg, and Michael Tomasello. 2011. "Collaboration Encourages Equal Sharing in Children but Not in Chimpanzees." *Nature* 476 (7360): 328–31. https://doi.org/10.1038/nature10278.

Hanish, Laura D., and Philip C. Rodkin. 2007. "Bridging Children's Social Development and Social Network Analysis." *New Directions for Child and Adolescent Development* 2007 (118): 1–8. https://doi.org/10.1002/cd.197.

Hanna, Elizabeth, and Andrew N. Meltzoff. 1993. "Peer Imitation by Toddlers in Laboratory, Home, and Day-Care Contexts: Implications for Social Learning and Memory." *Developmental Psychology* 29 (4): 701–10. https://doi.org/10.1037/0012-1649.29.4.701.

Hanus, Daniel, Natacha Mendes, Claudio Tennie, and Josep Call. 2011. "Comparing the Performances of Apes (Gorilla Gorilla, Pan Troglodytes, Pongo Pygmaeus) and Human Children (Homo Sapiens) in the Floating Peanut Task." Edited by Sarah Frances Brosnan. *PLoS ONE* 6 (6): e19555. https://doi.org/10.1371/journal.pone.0019555.

Harris, Paul L. 2012. *Trusting What You're Told: How Children Learn from Others*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Haun, Daniel B. M., Yvonne Rekers, and Michael Tomasello. 2014. "Children Conform to the Behavior of Peers; Other Great Apes Stick With What They Know." *Psychological Science* 25 (12): 2160–67. https://doi.org/10.1177/0956797614553235.

Haun, Daniel B. M., and Michael Tomasello. 2011. "Conformity to Peer Pressure in Preschool Children: Peer Pressure in Preschool Children." *Child Development* 82 (6): 1759–67. https://doi.org/10.1111/j.1467-8624.2011.01666.x.

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. "The Weirdest People in the World?" *Behavioral and Brain Sciences* 33 (2–3): 61–83. https://doi.org/10.1017/S0140525X0999152X.

Herrmann, Esther, Josep Call, María Victoria Hernàndez-Lloreda, Brian Hare, and Michael Tomasello. 2007. "Humans Have Evolved Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis." *Science* 317 (5843): 1360–66. https://doi.org/10.1126/science.1146282.

Hewlett, Barry S. 2016. "Social Learning and Innovation in Hunter- Gatherers." In *Social Learning and Innovation in Contemporary Hunter-Gatherers*, edited by H. Terrashima and Barry S. Hewlett, 1–15. Springer Japan.

Hewlett, Barry S., and Adam H. Boyette. 2012. "Play in Hunter-Gatherers." In *Evolution, Early Experience and Human Development.*, edited by D. Narvaez, J. Panksepp, A. Schore, and T. Gleason, 388–93. New York: Oxford University Press.

Heyes, C. M. 1994. "SOCIAL LEARNING IN ANIMALS: CATEGORIES AND MECHANISMS." *Biological Reviews* 69 (2): 207–31. https://doi.org/10.1111/j.1469-185X.1994.tb01506.x.

Hill, K. 2010. "Experimental Studies of Animal Social Learning in the Wild: Trying to Untangle the Mystery of Human Culture." *Learning & Behavior* 38 (3): 319–28. https://doi.org/10.3758/LB.38.3.319.

Hoffmann, Jessica, and Sandra Russ. 2012. "Pretend Play, Creativity, and Emotion Regulation in Children." *Psychology of Aesthetics, Creativity, and the Arts* 6 (2): 175–84. https://doi.org/10.1037/a0026299.

Hoicka, Elena, and Nameera Akhtar. 2011. "Preschoolers Joke with Jokers, but Correct Foreigners: Preschoolers Joke with Jokers, but Correct Foreigners." *Developmental Science* 14 (4): 848–58. https://doi.org/10.1111/j.1467-7687.2010.01033.x.

Hoicka, Elena, and Simone Bijvoet-van den Berg. 2013. "The Unusual Box Test: A Non-Verbal, Non-Representational Divergent Thinking Test for Toddlers." In , 6.

Hoicka, Elena, and Jessica Butcher. 2016. "Parents Produce Explicit Cues That Help Toddlers Distinguish Joking and Pretending." *Cognitive Science* 40 (4): 941–71. https://doi.org/10.1111/cogs.12264.

Hoicka, Elena, Rachael Mowat, Joanne Kirkwood, Tiffany Kerr, Megan Carberry, and Simone Bijvoet-van den Berg. 2016. "One-Year-Olds Think Creatively, Just Like Their Parents." *Child Development* 87 (4): 1099–1105. https://doi.org/10.1111/cdev.12531.

Hoicka, Elena, Stephanie Powell, Jenny Knight, and Megan Norwood. 2018. "Two-Year-Olds Can Socially Learn to Think Divergently." *British Journal of Developmental Psychology* 36 (1): 22–36. https://doi.org/10.1111/bjdp.12199.

Hopper, Lydia M. 2010. "'Ghost' Experiments and the Dissection of Social Learning in Humans and Animals." *Biological Reviews* 85 (4): 685–701. https://doi.org/10.1111/j.1469-185X.2010.00120.x.

Hopper, Lydia M., Emma G. Flynn, Lara A. N. Wood, and Andrew Whiten. 2010. "Observational Learning of Tool Use in Children: Investigating Cultural Spread through Diffusion Chains and Learning Mechanisms through Ghost Displays." *Journal of Experimental Child Psychology* 106 (1): 82–97. https://doi.org/10.1016/j.jecp.2009.12.001.

Hoppitt, William J.E., and Kevin N. Laland. 2013. *Social Learning: An Introduction to Mechanisms, Methods, and Models.* 1st ed. Princeton University Press.

Horner, Victoria, and Andrew Whiten. 2005. "Causal Knowledge and Imitation/Emulation Switching in Chimpanzees (Pan Troglodytes) and Children (Homo Sapiens)." *Animal Cognition* 8 (3): 164–81. https://doi.org/10.1007/s10071-004-0239-6.

Horner, Victoria, Andrew Whiten, Emma Flynn, and Frans B. M. de Waal. 2006. "Faithful Replication of Foraging Techniques along Cultural Transmission Chains by Chimpanzees and Children." *Proceedings of the National Academy of Sciences* 103 (37): 13878–83. https://doi.org/10.1073/pnas.0606015103.

House, Bailey R., Joan B. Silk, Joseph Henrich, H. Clark Barrett, Brooke A. Scelza, Adam H. Boyette, Barry S. Hewlett, Richard McElreath, and Stephen Laurence. 2013. "Ontogeny of Prosocial Behavior across Diverse Societies." *Proceedings of the National Academy of Sciences* 110 (36): 14586–91. https://doi.org/10.1073/pnas.1221217110.

Ivcevic, Zorana. 2009. "Creativity Map: Toward the next Generation of Theories of Creativity." *Psychology of Aesthetics, Creativity, and the Arts* 3 (1): 17–21. https://doi.org/10.1037/a0014918.

Jamieson, Susan. 2004. "Likert Scales: How to (Ab)Use Them." *Medical Education* 38 (12): 1217–18. https://doi.org/10.1111/j.1365-2929.2004.02012.x.

Jaswal, Vikram K., and Leslie A. Neely. 2006. "Adults Don't Always Know Best: Preschoolers Use Past Reliability Over Age When Learning New Words." *Psychological Science* 17 (9): 757–58. https://doi.org/10.1111/j.1467-9280.2006.01778.x.

Jipson, David M., and Jennifer L. Sobel. 2015. *Cognitive Development in Museum Settings: Relating Research and Practice*. 1st ed. Routledge. https://www.book2look.com/book/cirST7MmBl.

Kaufman, James C., and Ronald A. Beghetto. 2009. "Beyond Big and Little: The Four c Model of Creativity." *Review of General Psychology* 13 (1): 1–12. https://doi.org/10.1037/a0013688.

Kendal, Jeremy R. 2011. "Cultural Niche Construction and Human Learning Environments: Investigating Sociocultural Perspectives." *Biological Theory* 6 (3): 241–50. https://doi.org/10.1007/s13752-012-0038-2.

Kendal, Rachel L., Neeltje J. Boogert, Luke Rendell, Kevin N. Laland, Mike Webster, and Patricia L. Jones. 2018. "Social Learning Strategies: Bridge-Building between Fields." *Trends in Cognitive Sciences* 22 (7): 651–65. https://doi.org/10.1016/j.tics.2018.04.003.

Kendal, Rachel L., Bennett G. Galef, and Carel P. Van Schaik. 2010. "Social Learning Research Outside the Laboratory: How and Why?" *Learning & Behavior* 38 (3): 187–94. https://doi.org/10.3758/LB.38.3.187.

Kendal, Rachel L., Jeremy R. Kendal, Zarja Mursic, Claire Bailey-Ross, Hannah Rudman, Andy Lloyd, and Bethan Ross. 2016. "Designing for Creativity and Innovation in Informal Science Learning." *Informal Learning Review* 137: 20–24.

Kesek, Amanda, William A. Cunningham, Dominic J. Packer, and Philip David Zelazo. 2011. "Indirect Goal Priming Is More Powerful than Explicit Instruction in Children." *Developmental Science* 14 (5): 944–48. https://doi.org/10.1111/j.1467-7687.2011.01043.x.

Keupp, Stefanie, Tanya Behne, and Hannes Rakoczy. 2013. "Why Do Children Overimitate? Normativity Is Crucial." *Journal of Experimental Child Psychology* 116 (2): 392–406. https://doi.org/10.1016/j.jecp.2013.07.002.

Keupp, Stefanie, Tanya Behne, Joanna Zachow, Alina Kasbohm, and Hannes Rakoczy. 2015. "Over-Imitation Is Not Automatic: Context Sensitivity in Children's Overimitation and Action Interpretation of Causally Irrelevant Actions." *Journal of Experimental Child Psychology* 130 (February): 163–75. https://doi.org/10.1016/j.jecp.2014.10.005.

Klahr, D., and M. Nigam. 2004. "The Equivalence of Learning Paths in Early Science Instruction: Effects of Direct Instruction and Discovery Learning." *Psychological Science* 15 (10): 661–67. https://doi.org/10.1111/j.0956-7976.2004.00737.x.

Kline, Michelle Ann. 2014. "How to Learn about Teaching: An Evolutionary Framework for the Study of Teaching Behavior in Humans and Other Animals." *Behavioral and Brain Sciences*, May, 1–70. https://doi.org/10.1017/S0140525X14000090.

Köymen, Bahar, Elena Lieven, Denis A. Engemann, Hannes Rakoczy, Felix Warneken, and Michael Tomasello. 2014. "Children's Norm Enforcement in Their Interactions With Peers." *Child Development* 85 (3): 1108–22. https://doi.org/10.1111/cdev.12178.

Kruschke, John K., and Torrin M. Liddell. 2018. "Bayesian Data Analysis for Newcomers." *Psychonomic Bulletin & Review* 25 (1): 155–77. https://doi.org/10.3758/s13423-017-1272-1.

Kuhn, Deanna. 1989. “Children and Adults as Intuitive Scientists.” *Psychological Review* 96 (4): 674–89. http://dx.doi.org/10.1037/0033-295X.96.4.674.

Laland, Kevin N. 2004. “Social Learning Strategies.” *Animal Learning & Behavior* 32 (1): 4–14. https://doi.org/10.3758/BF03196002.

Lancy, David F. 2017. “Homo Faber Juvenalis: A Multidisciplinary Survey of Children as Tool Makers/Users.” *Childhood in the Past* 10 (1): 72–90. https://doi.org/10.1080/17585716.2017.1316010.

Leeuw, Joshua R. de. 2015. “JsPsych: A JavaScript Library for Creating Behavioral Experiments in a Web Browser.” *Behavior Research Methods* 47 (1): 1–12. https://doi.org/10.3758/s13428-014-0458-y.

Legare, Cristine H. 2017. “Cumulative Cultural Learning: Development and Diversity.” *Proceedings of the National Academy of Sciences* 114 (30): 7877–83. https://doi.org/10.1073/pnas.1620743114.

Legare, Cristine H., Robin Gose, and Cybil Guess. 2015. “Examining the Development of Scientific Reasoning in Context.” In *Cognitive Development in Museum Settings: Relating Research and Practice*, 1st ed., 138–48.

Legare, Cristine H., and Paul L. Harris. 2016. “The Ontogeny of Cultural Learning.” *Child Development* 87 (3): 633–42. https://doi.org/10.1111/cdev.12542.

Legare, Cristine H., and Mark Nielsen. 2015. “Imitation and Innovation: The Dual Engines of Cultural Learning.” *Trends in Cognitive Sciences* 19 (11): 688–99. https://doi.org/10.1016/j.tics.2015.08.005.

Legare, Cristine H., David M. Sobel, and Maureen Callanan. 2017. “Causal Learning Is Collaborative: Examining Explanation and Exploration in Social Contexts.” *Psychonomic Bulletin & Review* 24 (5): 1548–54. https://doi.org/10.3758/s13423-017-1351-3.

Liddell, Torrin M, and John K Kruschke. 2018. “Analyzing Ordinal Data with Metric Models: What Could Possibly Go Wrong?” *Open Science Framework*.

Lillard, Angeline S. 2017. “Why Do the Children (Pretend) Play?” *Trends in Cognitive Sciences* 21 (11): 826–34. https://doi.org/10.1016/j.tics.2017.08.001.

Lillard, Angeline S., Matthew D. Lerner, Emily J. Hopkins, Rebecca A. Dore, Eric D. Smith, and Carolyn M. Palmquist. 2013. “The Impact of Pretend Play on Children’s Development: A Review of the Evidence.” *Psychological Bulletin* 139 (1): 1–34. https://doi.org/10.1037/a0029321.

Lyons, Derek E., Diana H. Damrosch, Jennifer K. Lin, Deanna M. Macris, and Frank C. Keil. 2011. “The Scope and Limits of Overimitation in the Transmission of Artefact Culture.” *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 366 (1567): 1158–67. https://doi.org/10.1098/rstb.2010.0335.

Lyons, Derek E., Andrew G. Young, and Frank C. Keil. 2007. “The Hidden Structure of Overimitation.” *Proceedings of the National Academy of Sciences* 104 (50): 19751–56. https://doi.org/10.1073/pnas.0704452104.

Macdonald, C., and A. Whiten. 2011. “The ‘Living Links to Human Evolution’ Research Centre in Edinburgh Zoo: A New Endeavour in Collaboration.” *International Zoo Yearbook* 45 (1): 7–17. https://doi.org/10.1111/j.1748-1090.2010.00120.x.

Martin, Paul, and Paul Patrick Gordon Bateson. 2015. *Measuring Behaviour: An Introductory Guide*. 12th printing. Cambridge: Cambridge University Press.

McElreath, Richard. 2016. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. First edition. CRC Press.

McGraw, John J., Sebastian Wallot, Panagiotis Mitkidis, and Andreas Roepstorff. 2014. "Culture's Building Blocks: Investigating Cultural Evolution in a LEGO Construction Task." *Frontiers in Psychology* 5 (September). https://doi.org/10.3389/fpsyg.2014.01017.

McGuigan, Nicola, Emily Burdett, Vanessa Burgess, Lewis Dean, Amanda Lucas, Gillian Vale, and Andrew Whiten. 2017. "Innovation and Social Transmission in Experimental Micro-Societies: Exploring the Scope of Cumulative Culture in Young Children." *Phil. Trans. R. Soc. B* 372 (1735): 20160425. https://doi.org/10.1098/rstb.2016.0425.

McGuigan, Nicola, and Marcus Cubillo. 2013. "Information Transmission in Young Children: When Social Information Is More Important Than Nonsocial Information." *The Journal of Genetic Psychology* 174 (6): 605–19. https://doi.org/10.1080/00221325.2012.749833.

McGuigan, Nicola, and Murray Graham. 2010. "Cultural Transmission of Irrelevant Tool Actions in Diffusion Chains of 3- and 5-Year-Old Children." *European Journal of Developmental Psychology* 7 (5): 561–77. https://doi.org/10.1080/17405620902858125.

McGuigan, Nicola, Jenny Makinson, and Andrew Whiten. 2011. "From Over-Imitation to Super-Copying: Adults Imitate Causally Irrelevant Aspects of Tool Use with Higher Fidelity than Young Children." *British Journal of Psychology* 102 (1): 1–18. https://doi.org/10.1348/000712610X493115.

McGuigan, Nicola, and Stacey Robertson. 2015. "The Influence of Peers on the Tendency of 3- and 4-Year-Old Children to over-Imitate." *Journal of Experimental Child Psychology* 136 (August): 42–54. https://doi.org/10.1016/j.jecp.2015.03.004.

McGuigan, Nicola, Andrew Whiten, Emma Flynn, and Victoria Horner. 2007. "Imitation of Causally Opaque versus Causally Transparent Tool Use by 3- and 5-Year-Old Children." *Cognitive Development* 22 (3): 353–64. https://doi.org/10.1016/j.cogdev.2007.01.001.

Meltzoff, Andrew N. 1985. "Immediate and Deferred Imitation in Fourteen- and Twenty-Four-Month-Old Infants." *Child Development* 56 (1): 62–72. https://doi.org/10.2307/1130174.

———. 1988. "Infant Imitation after a 1-Week Delay: Long-Term Memory for Novel Acts and Multiple Stimuli." *Developmental Psychology* 24 (4): 470–76. https://doi.org/10.1037/0012-1649.24.4.470.

———. 1995. "Understanding the Intentions of Others: Re-Enactment of Intended Acts by 18-Month-Old Children." *Developmental Psychology* 31 (5): 838–50. https://doi.org/10.1037/0012-1649.31.5.838.

Meltzoff, Andrew N., and M. Keith Moore. 1983. "Newborn Infants Imitate Adult Facial Gestures." *Child Development* 54 (3): 702–9. https://doi.org/10.2307/1130058.

Mendes, Natacha, Daniel Hanus, and Josep Call. 2007. "Raising the Level: Orangutans Use Water as a Tool." *Biology Letters* 3 (5): 453–55. https://doi.org/10.1098/rsbl.2007.0198.

Mesoudi, A., and A. Whiten. 2008. "The Multiple Roles of Cultural Transmission Experiments in Understanding Human Cultural Evolution." *Philosophical Transactions of the Royal Society B: Biological Sciences* 363 (1509): 3489–3501. https://doi.org/10.1098/rstb.2008.0129.

Mesoudi, Alex. 2011. *Cultural Evolution*. Chicago: The University of Chicago Press. https://www.press.uchicago.edu/ucp/books/book/chicago/C/bo8787504.html.

Mesoudi, Alex, Lei Chang, Sasha R. X. Dall, and Alex Thornton. 2016. "The Evolution of Individual and Cultural Variation in Social Learning." *Trends in Ecology & Evolution* 31 (3): 215–25. https://doi.org/10.1016/j.tree.2015.12.012.

Miton, Helena, and Mathieu Charbonneau. 2018. "Cumulative Culture in the Laboratory: Methodological and Theoretical Challenges." *Proceedings of the Royal Society B: Biological Sciences* 285 (1879): 20180677. https://doi.org/10.1098/rspb.2018.0677.

Moll, H., and M. Tomasello. 2007. "Cooperation and Human Cognition: The Vygotskian Intelligence Hypothesis." *Philosophical Transactions of the Royal Society B: Biological Sciences* 362 (1480): 639–48. https://doi.org/10.1098/rstb.2006.2000.

Moraru, Cristina-Andreea, Juan-Carlos Gomez, and Nicola McGuigan. 2016. "Developmental Changes in the Influence of Conventional and Instrumental Cues on Over-Imitation in 3- to 6-Year-Old Children." *Journal of Experimental Child Psychology* 145 (May): 34–47. https://doi.org/10.1016/j.jecp.2015.11.017.

Moreau, C. Page, and Marit Gundersen Engeset. 2015. "The Downstream Consequences of Problem-Solving Mindsets: How Playing with LEGO Influences Creativity." *Journal of Marketing Research* 53 (1): 18–30. https://doi.org/10.1509/jmr.13.0499.

Morgan, Thomas J.H., Kevin N. Laland, and Paul L. Harris. 2015. "The Development of Adaptive Conformity in Young Children: Effects of Uncertainty and Consensus." *Developmental Science* 18 (4): 511–24. https://doi.org/10.1111/desc.12231.

Muthukrishna, Michael, and Joseph Henrich. 2016. "Innovation in the Collective Brain." *Phil. Trans. R. Soc. B* 371 (1690): 20150192. https://doi.org/10.1098/rstb.2015.0192.

Neldner, Karri, Ilana Mushin, and Mark Nielsen. 2017. "Young Children's Tool Innovation across Culture: Affordance Visibility Matters." *Cognition* 168 (November): 335–43. https://doi.org/10.1016/j.cognition.2017.07.015.

Newman, Sharlene D., Mitchell T. Hansen, and Arianna Gutierrez. 2016. "An FMRI Study of the Impact of Block Building and Board Games on Spatial Ability." *Frontiers in Psychology* 7 (August). https://doi.org/10.3389/fpsyg.2016.01278.

Nielsen, Mark. 2006. "Copying Actions and Copying Outcomes: Social Learning through the Second Year." *Developmental Psychology* 42 (3): 555–65. https://doi.org/10.1037/0012-1649.42.3.555.

———. 2013. “Young Children’s Imitative and Innovative Behaviour on the Floating Object Task: Children’s Imitative and Innovative Behaviour on the Floating Object Task.” *Infant and Child Development* 22 (1): 44–52. https://doi.org/10.1002/icd.1765.

Nielsen, Mark, Jessica Cucchiaro, and Jumana Mohamedally. 2012. “When the Transmission of Culture Is Child’s Play.” Edited by Alex Mesoudi. *PLoS ONE* 7 (3): e34066. https://doi.org/10.1371/journal.pone.0034066.

Nielsen, Mark, Chris Moore, and Jumana Mohamedally. 2012. “Young Children Overimitate in Third-Party Contexts.” *Journal of Experimental Child Psychology* 112 (1): 73–83. https://doi.org/10.1016/j.jecp.2012.01.001.

Nielsen, Mark, Ilana Mushin, Keyan Tomaselli, and Andrew Whiten. 2016. “Imitation, Collaboration, and Their Interaction Among Western and Indigenous Australian Preschool Children.” *Child Development* 87 (3): 795–806. https://doi.org/10.1111/cdev.12504.

Nielsen, Mark, and Keyan Tomaselli. 2010. “Overimitation in Kalahari Bushman Children and the Origins of Human Cultural Cognition.” *Psychological Science* 21 (5): 729–36. https://doi.org/10.1177/0956797610368808.

Nielsen, Mark, Keyan Tomaselli, Ilana Mushin, and Andrew Whiten. 2014. “Exploring Tool Innovation: A Comparison of Western and Bushman Children.” *Journal of Experimental Child Psychology* 126 (October): 384–94. https://doi.org/10.1016/j.jecp.2014.05.008.

Olson, Kristina R., and Elizabeth S. Spelke. 2008. “Foundations of Cooperation in Young Children.” *Cognition* 108 (1): 222–31. https://doi.org/10.1016/j.cognition.2007.12.003.

Over, Harriet, and Malinda Carpenter. 2012. “Putting the Social into Social Learning: Explaining Both Selectivity and Fidelity in Children’s Copying Behavior.” *Journal of Comparative Psychology* 126 (2): 182–92. https://doi.org/10.1037/a0024555.

———. 2013. “The Social Side of Imitation.” *Child Development Perspectives* 7 (1): 6–11. https://doi.org/10.1111/cdep.12006.

Over, Harriet, Adam Eggleston, Jenny Bell, and Yarrow Dunham. 2018. “Young Children Seek out Biased Information about Social Groups.” *Developmental Science* 21 (3): e12580. https://doi.org/10.1111/desc.12580.

Palan, Stefan, and Christian Schitter. 2018. “Prolific.Ac—A Subject Pool for Online Experiments.” *Journal of Behavioral and Experimental Finance* 17 (March): 22–27. https://doi.org/10.1016/j.jbef.2017.12.004.

Paradise, Ruth, and Barbara Rogoff. 2009. “Side by Side: Learning by Observing and Pitching In.” *Ethos* 37 (1): 102–38. https://doi.org/10.1111/j.1548-1352.2009.01033.x.

Piccuito, E., and P. Carruthers. 2014. “The Origins of Creativity.” In *The Philosophy of Creativity*, edited by E. S. Paul and S. B. Kaufman, 199–223. Oxford, UK: Oxford University Press. 10.1093/acprof:oso/9780199836963.001.0001.

Piffer, Davide. 2012. “Can Creativity Be Measured? An Attempt to Clarify the Notion of Creativity and General Directions for Future Research.” *Thinking Skills and Creativity* 7 (3): 258–64. https://doi.org/10.1016/j.tsc.2012.04.009.

Pinkham, Ashley M., and Vikram K. Jaswal. 2011. "Watch and Learn? Infants Privilege Efficiency Over Pedagogy During Imitative Learning: EFFICIENCY AND PEDAGOGY IN IMITATIVE LEARNING." *Infancy* 16 (5): 535–44. https://doi.org/10.1111/j.1532-7078.2010.00059.x.

R Core Team. 2013. *R: A Language and Environment for Statistical Computing.* (version 3.3.2). Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Rakoczy, Hannes, Nina Brosche, Felix Warneken, and Michael Tomasello. 2009. "Young Children's Understanding of the Context-Relativity of Normative Rules in Conventional Games." *British Journal of Developmental Psychology* 27 (2): 445–56. https://doi.org/10.1348/026151008X337752.

Rakoczy, Hannes., Katharina. Hamann, Felix. Warneken, and Michael. Tomasello. 2010. "Bigger Knows Better: Young Children Selectively Learn Rule Games from Adults Rather than from Peers." *British Journal of Developmental Psychology* 28 (4): 785–98. https://doi.org/10.1348/026151009X479178.

Ramsey, Grant, Meredith L. Bastian, and Carel van Schaik. 2007. "Animal Innovation Defined and Operationalized." *Behavioral and Brain Sciences* 30 (04). https://doi.org/10.1017/S0140525X07002373.

Rawlings, Bruce, Emma Flynn, and Rachel Kendal. 2017. "To Copy or To Innovate? The Role of Personality and Social Networks in Children's Learning Strategies." *Child Development Perspectives* 11 (1): 39–44. https://doi.org/10.1111/cdep.12206.

Reader, Simon M., and Kevin N. Laland. 2003. "Animal Innovation: An Introduction." In *Animal Innovation*, 3–35. New York, NY, US: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198526223.003.0001.

Reifel, Stuart. 1984. "Block Construction: Children's Developmental Landmarks in Representation of Space." *Young Children* 40 (1): 61–67.

Reindl, E., I. A. Apperly, S. R. Beck, and C. Tennie. 2017. "Young Children Copy Cumulative Technological Design in the Absence of Action Information." *Scientific Reports* 7 (1). https://doi.org/10.1038/s41598-017-01715-2.

Reindl, E., S. R. Beck, I. A. Apperly, and C. Tennie. 2016. "Young Children Spontaneously Invent Wild Great Apes' Tool-Use Behaviours." *Proc. R. Soc. B* 283 (1825): 20152402. https://doi.org/10.1098/rspb.2015.2402.

Reindl, Eva, and Claudio Tennie. 2018. "Young Children Fail to Generate an Additive Ratchet Effect in an Open-Ended Construction Task." Edited by Kenny Smith. *PLOS ONE* 13 (6): e0197828. https://doi.org/10.1371/journal.pone.0197828.

Rendell, Luke, Laurel Fogarty, William J. E. Hoppitt, Thomas J. H. Morgan, Mike M. Webster, and Kevin N. Laland. 2011. "Cognitive Culture: Theoretical and Empirical Insights into Social Learning Strategies." *Trends in Cognitive Sciences* 15 (2): 68–76. https://doi.org/10.1016/j.tics.2010.12.002.

Rudman, Hannah, Claire Bailey-Ross, Jeremy Kendal, Zarja Mursic, Andy Lloyd, Bethan Ross, and Rachel L. Kendal. 2017. "Multidisciplinary Exhibit Design in a Science Centre: A Participatory Action Research Approach." *Educational Action Research*, August, 1–22. https://doi.org/10.1080/09650792.2017.1360786.

Runco, Mark A. 1992. "Children's Divergent Thinking and Creative Ideation." *Developmental Review* 12 (3): 233–64. https://doi.org/10.1016/0273-2297(92)90010-Y.

Runco, Mark A. 2014. "'Big C, Little c' Creativity as a False Dichotomy: Reality Is Not Categorical." *Creativity Research Journal* 26 (1): 131–32. https://doi.org/10.1080/10400419.2014.873676.

Runco, Mark A., and Selcuk Acar. 2012. "Divergent Thinking as an Indicator of Creative Potential." *Creativity Research Journal* 24 (1): 66–75. https://doi.org/10.1080/10400419.2012.652929.

Runco, Mark A., and Garrett J. Jaeger. 2012. "The Standard Definition of Creativity." *Creativity Research Journal* 24 (1): 92–96. https://doi.org/10.1080/10400419.2012.650092.

Rusher, Anne Spidell, David R. Cross, and Anne M. Ware. 1995. "Infant and Toddler Play: Assessment of Exploratory Style and Development Level." *Early Childhood Research Quarterly* 10 (3): 297–315. https://doi.org/10.1016/0885-2006(95)90009-8.

Ryalls, Brigette Oliver, Robina E. Gul, and Kenneth R. Ryalls. 2000. "Infant Imitation of Peer and Adult Models: Evidence for a Peer Model Advantage." *Merrill-Palmer Quarterly* 46 (1): 188–202. https://www.jstor.org/stable/23093348.

Sandifer, Cody. 2003. "Technological Novelty and Open-Endedness: Two Characteristics of Interactive Exhibits That Contribute to the Holding of Visitor Attention in a Science Museum." *Journal of Research in Science Teaching* 40 (2): 121–37. https://doi.org/10.1002/tea.10068.

Sawyer, R. Keith. 2012. *Explaining Creativity: The Science of Human Innovation*. 2nd ed. New York, New York: Oxford University Press.

Sawyer, R. Keith, Vera John-Steiner, Seana Moran, Robert J. Sternberg, David Henry Feldman, Jeanne Nakamura, and Mihaly Csikszentmihalyi. 2003. *Creativity and Development*. New York, New York: Oxford University Press.

Schijndel, Tessa J. P. Van, and Maartje E. J. Raijmakers. 2016. "Parent Explanation and Preschoolers' Exploratory Behavior and Learning in a Shadow Exhibition." *Science Education* 100 (1): 153–78. https://doi.org/10.1002/sce.21193.

Schijndel, Tessa J. P. van, Brenda R. J. Jansen, and Maartje E. J. Raijmakers. 2018. "Do Individual Differences in Children's Curiosity Relate to Their Inquiry-Based Learning?" *International Journal of Science Education* 40 (9): 996–1015. https://doi.org/10.1080/09500693.2018.1460772.

Schillaci, Rebecca Seston, and Deborah Kelemen. 2014. "Children's Conformity When Acquiring Novel Conventions: The Case of Artifacts." *Journal of Cognition and Development* 15 (4): 569–83. https://doi.org/10.1080/15248372.2013.784973.

Schoot, Rens van de, and Sarah Depaoli. 2014. "Bayesian Analyses: Where to Start and What to Report." *European Health Psychologist* 16 (2): 75–84.

Schoot, Rens van de, David Kaplan, Jaap Denissen, Jens B. Asendorpf, Franz J. Neyer, and Marcel A.G. van Aken. 2014. "A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research." *Child Development* 85 (3): 842–60. https://doi.org/10.1111/cdev.12169.

Seehagen, Sabine, and Jane S. Herbert. 2011. "Infant Imitation From Televised Peer and Adult Models." *Infancy* 16 (2): 113–36. https://doi.org/10.1111/j.1532-7078.2010.00045.x.

Sheridan, Kimberly M., Abigail W. Konopasky, Sophie Kirkwood, and Margaret A. Defeyter. 2016. "The Effects of Environment and Ownership on Children's Innovation of Tools and Tool Material Selection." *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1690): 20150191. https://doi.org/10.1098/rstb.2015.0191.

Shneidman, Laura, Hyowon Gweon, Laura E. Schulz, and Amanda L. Woodward. 2016. "Learning From Others and Spontaneous Exploration: A Cross-Cultural Investigation." *Child Development* 87 (3): 723–35. https://doi.org/10.1111/cdev.12502.

Smaldino, Paul E., and Richard McElreath. 2016. "The Natural Selection of Bad Science." *Royal Society Open Science* 3 (9): 160384. https://doi.org/10.1098/rsos.160384.

Stan Development Team. 2018. *RStan: The R Interface to Stan* (version 2.17.3). http://mc-stan.org/.

Stannard, Laura, Charles H. Wolfgang, Ithel Jones, and Pamela Phelps. 2001. "A Longitudinal Study of the Predictive Relations Among Construction Play and Mathematical Achievement." *Early Child Development and Care* 167 (1): 115–25. https://doi.org/10.1080/0300443011670110.

Stein, Morris I. 1953. "Creativity and Culture." *The Journal of Psychology* 36 (2): 311–22. https://doi.org/10.1080/00223980.1953.9712897.

Sternberg, Robert J. 2006. "The Nature of Creativity." *Creativity Research Journal* 18 (1): 87–98. https://doi.org/10.1207/s15326934crj1801_10.

Symonds, Matthew R. E., and Adnan Moussalli. 2011. "A Brief Guide to Model Selection, Multimodel Inference and Model Averaging in Behavioural Ecology Using Akaike's Information Criterion." *Behavioral Ecology and Sociobiology* 65 (1): 13–21. https://doi.org/10.1007/s00265-010-1037-6.

Tennie, Claudio, Christine Caldwell, and Lewis G. Dean. 2018. "Culture, Cumulative." In *The International Encyclopedia of Anthropology*, 1–7. American Cancer Society. https://doi.org/10.1002/9781118924396.wbiea1998.

Tennie, Claudio, Josep Call, and Michael Tomasello. 2006. "Push or Pull: Imitation vs. Emulation in Great Apes and Human Children." *Ethology* 112 (12): 1159–69. https://doi.org/10.1111/j.1439-0310.2006.01269.x.

———. 2009. "Ratcheting up the Ratchet: On the Evolution of Cumulative Culture." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 364 (1528): 2405–15. https://doi.org/10.1098/rstb.2009.0052.

Tennie, Claudio, Victoria Walter, Anja Gampe, Malinda Carpenter, and Michael Tomasello. 2014. "Limitations to the Cultural Ratchet Effect in Young Children." *Journal of Experimental Child Psychology* 126 (October): 152–60. https://doi.org/10.1016/j.jecp.2014.04.006.

Terrashima, H., and Barry S. Hewlett. 2016. *Social Learning and Innovation in Contemporary Hunter-Gatherers*.

Tomasello, Michael. 1999. *The Cultural Origins of Human Cognition*. 4. print. Cambridge, Mass.: Harvard Univ. Press.

———. 2009. *Why We Cooperate: Based on the 2008 Tanner Lectures on Human Values at Stanford University*. Cambridge, Massachusetts: The MIT Press.

Tomasello, Michael, Ann Cale Kruger, and Hilary Horn Ratner. 1993. "Cultural Learning." *Behavioral and Brain Sciences* 16 (3): 495–511. https://doi.org/10.1017/S0140525X0003123X.

Turner, Cameron, Luc-Alain Giraldeau, and Emma Flynn. 2017. "How Does the Reliability of a Model Affect Children's Choice to Learn Socially or Individually?" *Evolution and Human Behavior* 38 (3): 341–49. https://doi.org/10.1016/j.evolhumbehav.2016.11.005.

Van Schijndel, Tessa J. P., Rooske K. Franse, and Maartje E. J. Raijmakers. 2010. "The Exploratory Behavior Scale: Assessing Young Visitors' Hands-on Behavior in Science Museums." *Science Education* 94 (5): 794–809. https://doi.org/10.1002/sce.20394.

Wallach, M. A., and N Kogan. 1965. *Modes of Thinking in Young Children: A Study of the Creativity-Intelligence Distinction*. Oxford, England: Holt, Rinehart and Winston.

Wang, Zhidan, Rebecca A. Williamson, and Andrew N. Meltzoff. 2015. "Imitation as a Mechanism in Cognitive Development: A Cross-Cultural Investigation of 4-Year-Old Children's Rule Learning." *Frontiers in Psychology* 6 (May). https://doi.org/10.3389/fpsyg.2015.00562.

Want, Stephen C., and Paul L. Harris. 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): 1–14. https://doi.org/10.1111/1467-7687.00194.

Warneken, F., and M. Tomasello. 2006. "Altruistic Helping in Human Infants and Young Chimpanzees." *Science* 311 (5765): 1301–3. https://doi.org/10.1126/science.1121448.

Warneken, Felix, Frances Chen, and Michael Tomasello. 2006. "Cooperative Activities in Young Children and Chimpanzees." *Child Development* 77 (3): 640–63. https://doi.org/10.1111/j.1467-8624.2006.00895.x.

Warneken, Felix, Karoline Lohse, Alicia P. Melis, and Michael Tomasello. 2011. "Young Children Share the Spoils After Collaboration." *Psychological Science* 22 (2): 267–73. https://doi.org/10.1177/0956797610395392.

Wasserstein, Ronald L., and Nicole A. Lazar. 2016. "The ASA's Statement on p-Values: Context, Process, and Purpose." *The American Statistician* 70 (2): 129–33. https://doi.org/10.1080/00031305.2016.1154108.

Watson, David, and Lee Anna Clark. 1997. "Extraversion and Its Positive Emotional Core." In *Handbook of Personality Psychology*, 767–93. Elsevier. https://doi.org/10.1016/B978-012134645-4/50030-5.

Weir, A. A. S., Jackie Chappell, and Alex Kacelnik. 2002. "Shaping of Hooks in New Caledonian Crows." *Science* 297 (5583): 981–981. https://doi.org/10.1126/science.1073433.

Whalley, Clare L., Nicola Cutting, and Sarah R. Beck. 2017. "The Effect of Prior Experience on Children's Tool Innovation." *Journal of Experimental Child Psychology* 161 (September): 81–94. https://doi.org/10.1016/j.jecp.2017.03.009.

Whiten, Andrew, Gillian Allan, Siobahn Devlin, Natalie Kseib, Nicola Raw, and Nicola McGuigan. 2016. "Social Learning in the Real-World: 'Over-Imitation' Occurs in Both Children and Adults Unaware of Participation in an Experiment and Independently of Social Interaction." Edited by Rachel L Kendal. *PLOS ONE* 11 (7): e0159920. https://doi.org/10.1371/journal.pone.0159920.

Whiten, Andrew, and Emma Flynn. 2010. "The Transmission and Evolution of Experimental Microcultures in Groups of Young Children." *Developmental Psychology* 46 (6): 1694–1709. https://doi.org/10.1037/a0020786.

Whiten, Andrew, Nicola McGuigan, Sarah Marshall-Pescini, and Lydia M. Hopper. 2009. "Emulation, Imitation, over-Imitation and the Scope of Culture for Child and Chimpanzee." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 364 (1528): 2417–28. https://doi.org/10.1098/rstb.2009.0069.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Second edition. Use R! Cham: Springer.

Wilks, Matti, Emma Collier-Baker, and Mark Nielsen. 2015. "Preschool Children Favor Copying a Successful Individual over an Unsuccessful Group." *Developmental Science* 18 (6): 1014–24. https://doi.org/10.1111/desc.12274.

Williamson, Rebecca A., Andrew N. Meltzoff, and Ellen M. Markman. 2008. "Prior Experiences and Perceived Efficacy Influence 3-Year-Olds' Imitation." *Developmental Psychology* 44 (1): 275–85. https://doi.org/10.1037/0012-1649.44.1.275.

Witmer, Jeff. 2017. "Bayes and MCMC for Undergraduates." *The American Statistician* 71 (3): 259–64. https://doi.org/10.1080/00031305.2017.1305289.

Wolfgang, Charles H., Laura L. Stannard, and Ithel Jones. 2001. "Block Play Performance Among Preschoolers As a Predictor of Later School Achievement in Mathematics." *Journal of Research in Childhood Education* 15 (2): 173–80. https://doi.org/10.1080/02568540109594958.

Wood, Lara A., Rachel A. Harrison, Amanda J. Lucas, Nicola McGuigan, Emily R.R. Burdett, and Andrew Whiten. 2016. "'Model Age-Based' and 'Copy When Uncertain' Biases in Children's Social Learning of a Novel Task." *Journal of Experimental Child Psychology* 150 (October): 272–84. https://doi.org/10.1016/j.jecp.2016.06.005.

Wood, Lara A., Rachel L. Kendal, and Emma G. Flynn. 2012. "Context-Dependent Model-Based Biases in Cultural Transmission: Children's Imitation Is Affected by Model Age over Model Knowledge State." *Evolution and Human Behavior* 33 (4): 387–94. https://doi.org/10.1016/j.evolhumbehav.2011.11.010.

———. 2013a. "Copy Me or Copy You? The Effect of Prior Experience on Social Learning." *Cognition* 127 (2): 203–13. https://doi.org/10.1016/j.cognition.2013.01.002.

———. 2013b. “Whom Do Children Copy? Model-Based Biases in Social Learning.” *Developmental Review* 33 (4): 341–56. https://doi.org/10.1016/j.dr.2013.08.002.

———. 2015. “Does a Peer Model’s Task Proficiency Influence Children’s Solution Choice and Innovation?” *Journal of Experimental Child Psychology* 139 (November): 190–202. https://doi.org/10.1016/j.jecp.2015.06.003.

Zajonc, Robert B. 1965. “Social Facilitation.” *Science* 149 (3681): 269–74. https://www.jstor.org/stable/1715944.

Zmyj, Norbert, Gisa Aschersleben, Wolfgang Prinz, and Moritz Daum. 2012. “The Peer Model Advantage in Infants’ Imitation of Familiar Gestures Performed by Differently Aged Models.” *Frontiers in Psychology* 3. https://doi.org/10.3389/fpsyg.2012.00252.

Zosh, Jennifer M., Kathy Hirsh-Pasek, Roberta Michnick Golinkoff, and Rebecca A. Dore. 2017. “Where Learning Meets Creativity: The Promise of Guided Play.” In *Creative Contradictions in Education*, 165–80. Creativity Theory and Action in Education. Springer, Cham. https://doi.org/10.1007/978-3-319-21924-0_10.

Appendix 1 - Consent form (Chapter 3)

Department of Anthropology,
Durham, DH1 3DE
Contact Name (researcher): Zarja Muršič
Tel. No.: 07534609896
Email: zarja.mursic@durham.ac.uk
Supervisor: Rachel Kendal
Tel. No.: +44 (0) 191 33 41627
Email: rachel.kendal@durham.ac.uk

Information and Consent

This is an information and consent form regarding your child's participation in a study that I [Zarja Muršič] am running at the Centre for Life.

The study is called '*Instructions vs. discovery learning: Interaction with exhibits within the Curiosity Zone of the Centre for Life, Newcastle*' (supervised by Dr Rachel Kendal, Durham University). The Curiosity Zone in the Centre for Life was developed to foster children's spontaneous exploration and development of scientific thinking. However, some adults have expressed a desire for instructions on exhibits. Therefore, with this study we intend to research how instructions affect children's interaction with exhibits. The main aim is to determine whether instructions are needed in the Curiosity Zone or not, and if needed, what type of instructions best facilitate the aims of enhancing creativity, exploration and scientific thinking in children.

Participation in the study will not take more than 20 minutes and has it been designed to be enjoyable for children taking part. Your child will interact with a task, which measures creativity. Afterwards s/he will be free to interact with the exhibit. When your child decides to stop investigating the exhibit, s/he will be asked a few questions and posed a new task, which measures spontaneous scientific thinking.

Your child can decide to withdraw from the study at any time, and for any reason. Also you may withdraw your child from the study. Child will receive a small reward (sticker or eraser) for their participation regardless of their performance.

Your child's interaction with the exhibit and subsequently presented task will be recorded on video. Also, the short interview will be audio/video recorded. The recordings will serve as a memory aid for me and will only be viewed/listened to by me. If you wish to view the footage, arrangements can be made for you to do this. I may wish to use the footage to illustrate my study's procedure and findings to other academics. You will be able to indicate (below) if you do not wish me to use any recordings in this way. I alone will know your child's behavioural responses and the data collected will be anonymised. All recordings will be destroyed at the end of the study.

I have a full Criminal Records check from Slovenia and have been approved for working with children. The study I am running has full ethical approval from Durham University as well as the Centre for Life. Finally, a report with the study's findings will be available at the Centre for life once the research has been completed.

I have met with, and fully briefed, the Director of Science Communication at the Centre for Life who has given me consent to work with visiting children. If you are willing for your child to participate in the study please complete the slip below. In addition, I would be glad to answer any questions you may have regarding this study.

Many thanks,

Zarja Muršič
PhD Candidate

Department of Anthropology, Durham University
Email: zarja.mursic@durham.ac.uk
Supervisor: Dr Rachel Kendal

Parent/Guardian, please complete the below:

Child's name:

Child's age:

Child's sex (please circle): Female Male

Does your child have any siblings? If so, how many of them are older/younger than him/her?
Number of older siblings……
Number of younger siblings …..

Have you visited the Curiosity Zone (of the Centre for Life) before? (Please circle)

Yes No

If yes, how many times approximately? ………

Please tick the appropriate box below:

☐ I **consent** to my child participating in the study (including video/audio recordings being taken).

☐ I **consent** to my child participating in the study (including video/audio recordings being taken). But **do NOT want recordings to be viewed** by other academics.

Signature: Date:

I wish to be sent a summary of the study's results when it is completed (please circle):

Yes No
(If yes, please provide email address)

Email:

Appendix 2 - Consent screen and consent application (Chapter 4)

Figure 57: Consent on a tablet computer embedded in the Interactive Research Pod

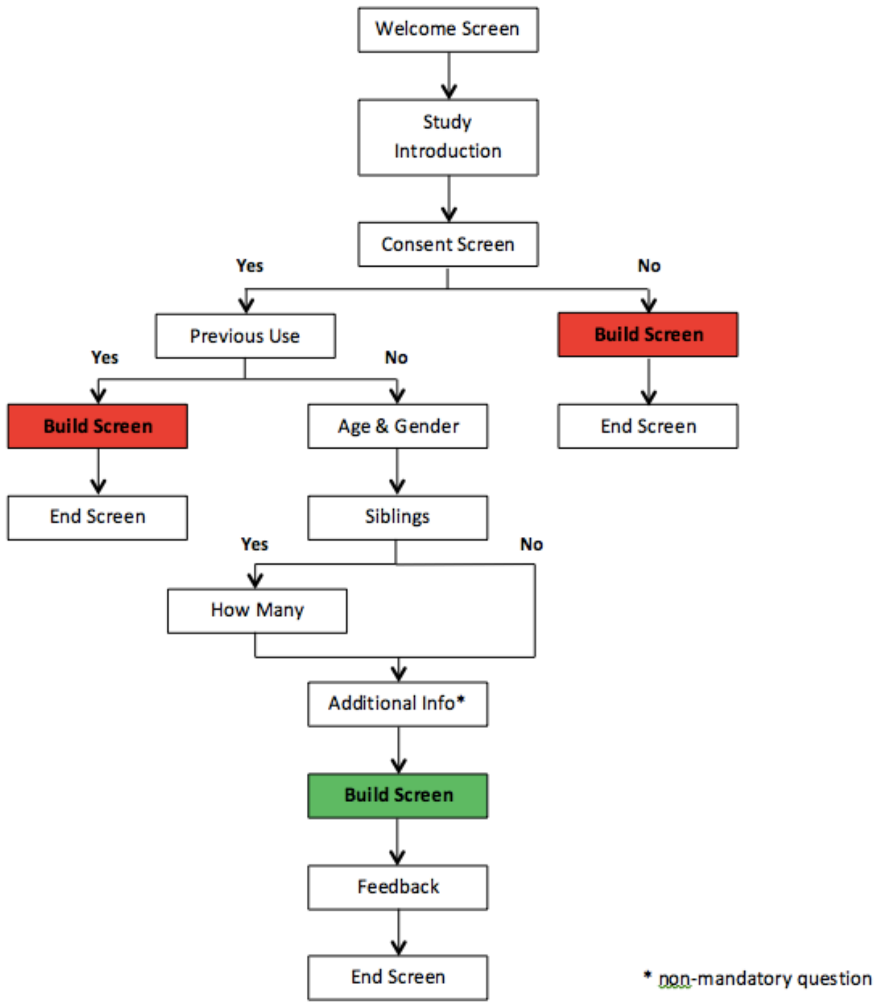


Figure 58: Questions included in the consent form at the Interactive research pod

Appendix 3 - Consent form (Chapter 5)

Department of Anthropology
Durham, DH1 3DE
Contact Name (researcher): Zarja Muršič
Tel. No.: 07534609896
Email: zarja.mursic@durham.ac.uk
Supervisor: Rachel Kendal
Tel. No.: +44 (0) 191 33 41627
Email: rachel.kendal@durham.ac.uk

Dear Parent/Guardian,

This is an information and consent form regarding your child's participation in a study that we [Zarja Muršič and Guy Lavender-Forsyth] are running at the Centre for Life.

The study is called '*Do children innovate in groups?*' (supervised by Dr Rachel Kendal & Dr Jeremy Kendal). Our main aim is to investigate whether children are better at innovating when working in a group as opposed to when working alone.

Your child's participation in the study would take approximately 10 minutes and would involve them building with wooden blocks, at a permanent exhibit in the new Brain Zone at the Centre for Life. The exhibit has been designed to be enjoyable for children taking part. At the end of the building phase, we will ask your child to name some of his/her best friends (we will not record the actual names given). This will enable me to investigate social factors that may play a role in group creativity. Your child will be rewarded for taking part with a sticker/small eraser etc. and can decide to leave the study, at any time, for any reason.

We will videotape children while building as the recordings will serve as a memory aid for me. Some recordings may be used to illustrate our study's procedure and findings to other academics. You can indicate (below) if you do not wish us to use the recordings in this way. We alone will know your child's identity and the data collected will be fully anonymised. If you wish to view the footage, arrangements can be made for you to do this. All recordings will be kept securely and destroyed at the end of the study.

We have a full DBS check and have been approved for working with children. We have worked with children in the past as researchers at the Centre for Life. The study has full ethical approval from Durham University as well as the Centre for Life. Finally, a report with the study’s findings will be available at the Centre for Life once the research has been completed.

If you are willing for your child to participate in the study please complete the slip below. In addition, we would be glad to answer any questions you may have regarding this study via email.

Many thanks,

Zarja Muršič (PhD Candidate)

Guy Lavender-Forsyth (Research assistant)

Parent/Guardian, please complete the below:

CHILD’S NAME:

CHILD’S DATE OF BIRTH (day/month/year):

CHILD’S SEX (please circle): FEMALE MALE

Does your child have any SIBLINGS? (please circle) YES NO

If so:

Number of YOUNGER _______

Number of SAME AGE _______

Number of OLDER _______

Have you visited the Centre for Life since Easter this year (2016)? (Please circle) YES

NO

Please tick the appropriate box below:

☐ Full consent: I **consent** to my child participating in the study (including video/audio recordings being taken).

or

☐ Partial consent: I **consent** to my child participating in the study (including video/audio recordings being taken). But **do NOT want recordings to be viewed** by other academics.

Signature: Date:

I wish to be sent a summary of the study's results when it is completed (please circle):

Yes No

(If yes, please provide email address)

Email:

Appendix 4 - Consent form from the web application (Chapters 3, 4, 5)

Consent form

I consent to participate in this session, which will involve 130 pairs of pictures I will have to rate by similarity on the Likert scale from 1 (very similar) to 7 (very different).

I understand that all data will be kept confidential by the researcher. My personal information will not be stored with the data. I am free to withdraw at any time without giving a reason.

I consent to the publication of study results as long as the information is anonymous so that no identification of participants can be made.

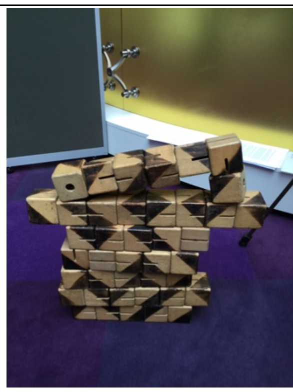
The study has received approval from the Research Ethics Committee by Department of Anthropology of the University of Durham.

☐ I have read and understand the explanations and I voluntarily consent to participate in this study.

Start

Appendix 5 – Examples of the structures children built (Chapters 3, 4, 5)

Table 35: Examples of the structures children built in the study presented in Chapter 3 where children were either exposed to direct instructions, scaffolding (open questions) or no instructions about how to use blocks prior to the building. Each row from left to right presents structures the same individual built. Under each image is the answer the child provided to the question of what did s/he built.

First structure	Second structure	Third structure
Condition: **Direct instruction**		
 diplodocus	 T-rex	 tower
 castle	 tower	 house
 garden for animals	 stairs	 towers


caterpillar


bridge


castle


bridge


house


castle

Condition: **Scaffolding (open questions)**


bridge


statue


obstacle course


towers


towers

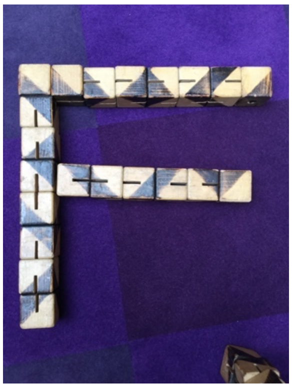

towers

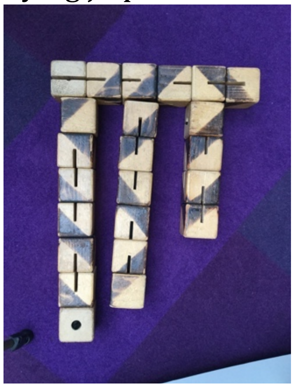

snowskater with legs


boat


flying jetpack


house


letter F


garden


nothing


tower


castle

Condition: **No instructions**


castle

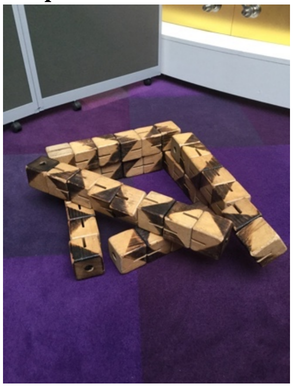

house


house


steps


climbing area


skateboarding ring


robot

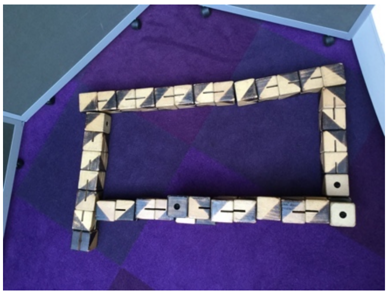

elephant

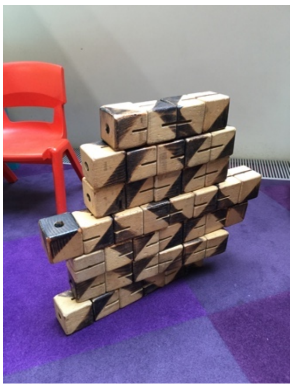

place for animals


statue


car

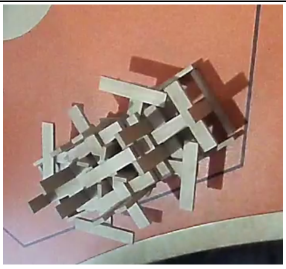

airplane


castle


cat


wall

Table 36: Examples of the structures children built in the study presented in Chapter 4 where children were either exposed to others building and could see them through transparent partitions (social condition) or could not see them due to the opaque partitions (asocial and asocial control condition). Each individual built only one structure and each image corresponds to one individual in the specific condition. Under each image is the answer the child provided to the question of what did s/he built.

Social condition (transparent partitions)		
 engine	 cinema	 house
 tank	 fort	
Asocial condition (opaque partitions, children in pairs built at similar time)		
 house	 castle	 human
 castle	 stadium	

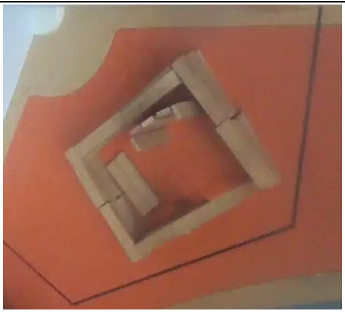
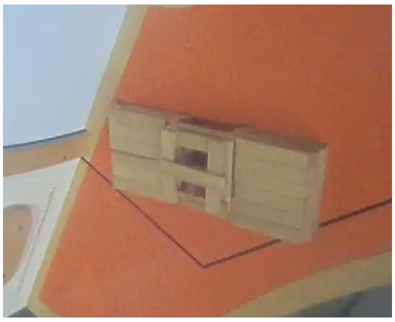
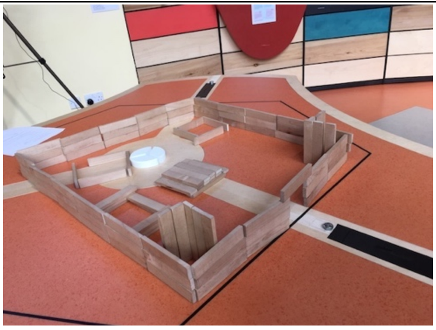
Asocial control condition (opaque partitions, children in pairs building alone at the IRP and pairs constituted of children who did not built on the same day)		
 house	 tower	 jenga
 tower	 basket	

Table 37: Examples of the structures children built in the study presented in Chapter 5 where children built in pairs and groups and could cooperate, collaborate, instruct, copy or discuss their building. Each pair or group built one structure. Under each image is the answer the pair or group provided to the question of what did they built.

Pairs and groups of children building together		
 tower	 castle	 human
 house	 tower	

Appendix 6 - Burn-in period for the online ratings of the originality of structures

In all three empirical chapters (Chapter 3, 4, 5) each adult rater rated between 110 to 124 pairs of images. Before they started rating, the participants were only exposed to one example. Therefore, I prepared three data sets excluding the first 10, 20 and 30 pairs that individuals rated and randomly selected smaller data frames (5000 ratings) to compare them to each other and see whether there was a burn-in period in rating the pairs of images. The models were intercept-only ordered categorical models predicting the ratings of the pairs with weakly informative priors. The models were approximated by a Gaussian distribution with the map function in the rethinking package (McElreath, 2016). There were no differences among estimates in smaller data frames which excluded some of the starting trials (Figures 59, 60, 61).

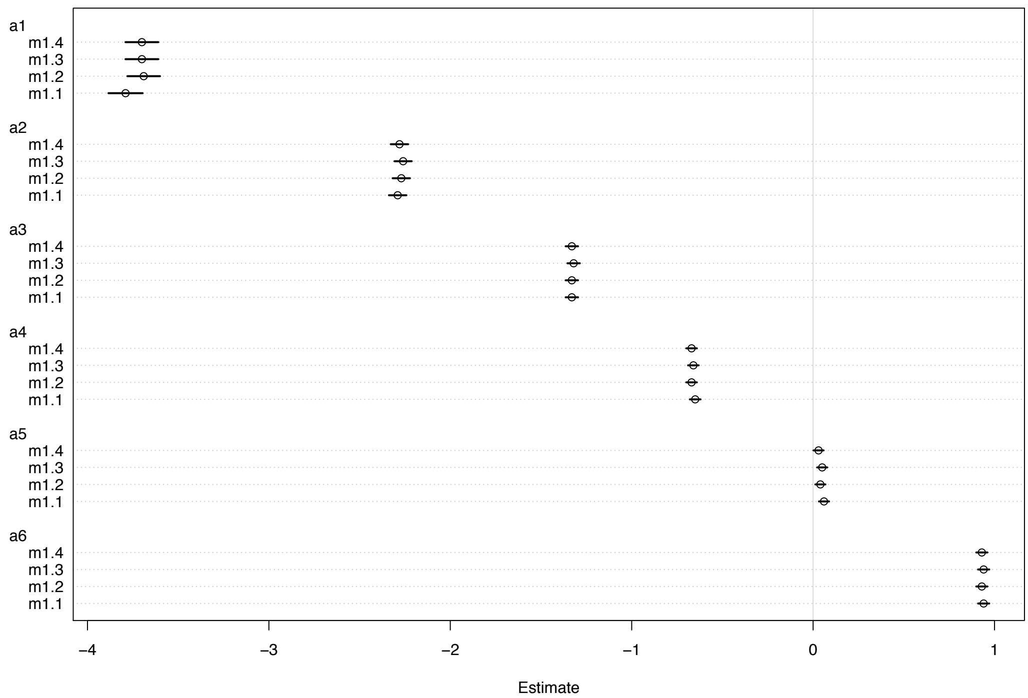


Figure 59: Estimates of the intercepts from Chapter 3 of ordinal categorical model with excluded pairs (without first 10, 20, 30 pairs) that individuals rated. a1, a2, a3, a4, a5, a6 correspond to the cutpoints for each value on the Likert scale from the ordinal categorical models.

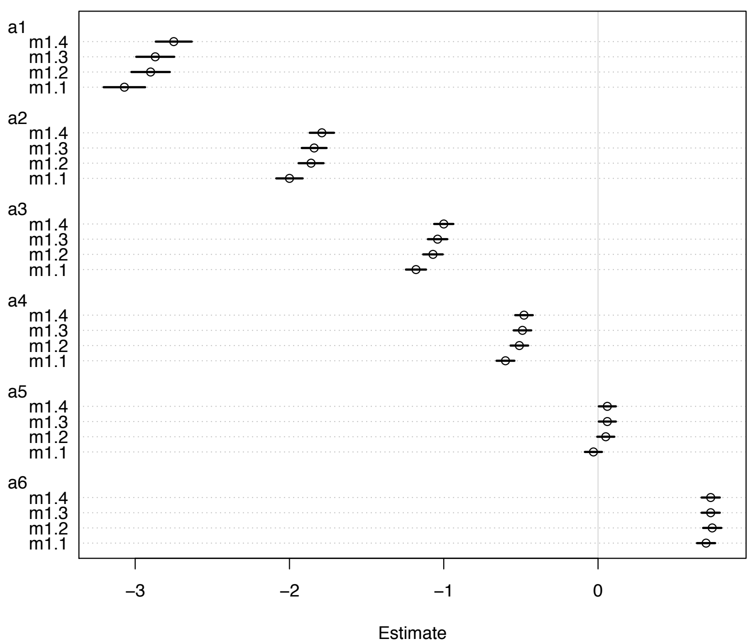


Figure 60: Estimates of the intercepts from Chapter 4 of ordinal categorical model with excluded pairs (without first 10, 20, 30 pairs) that individuals rated. a1, a2, a3, a4, a5, a6 correspond to the cutpoints for each value on the Likert scale from the ordinal categorical models.

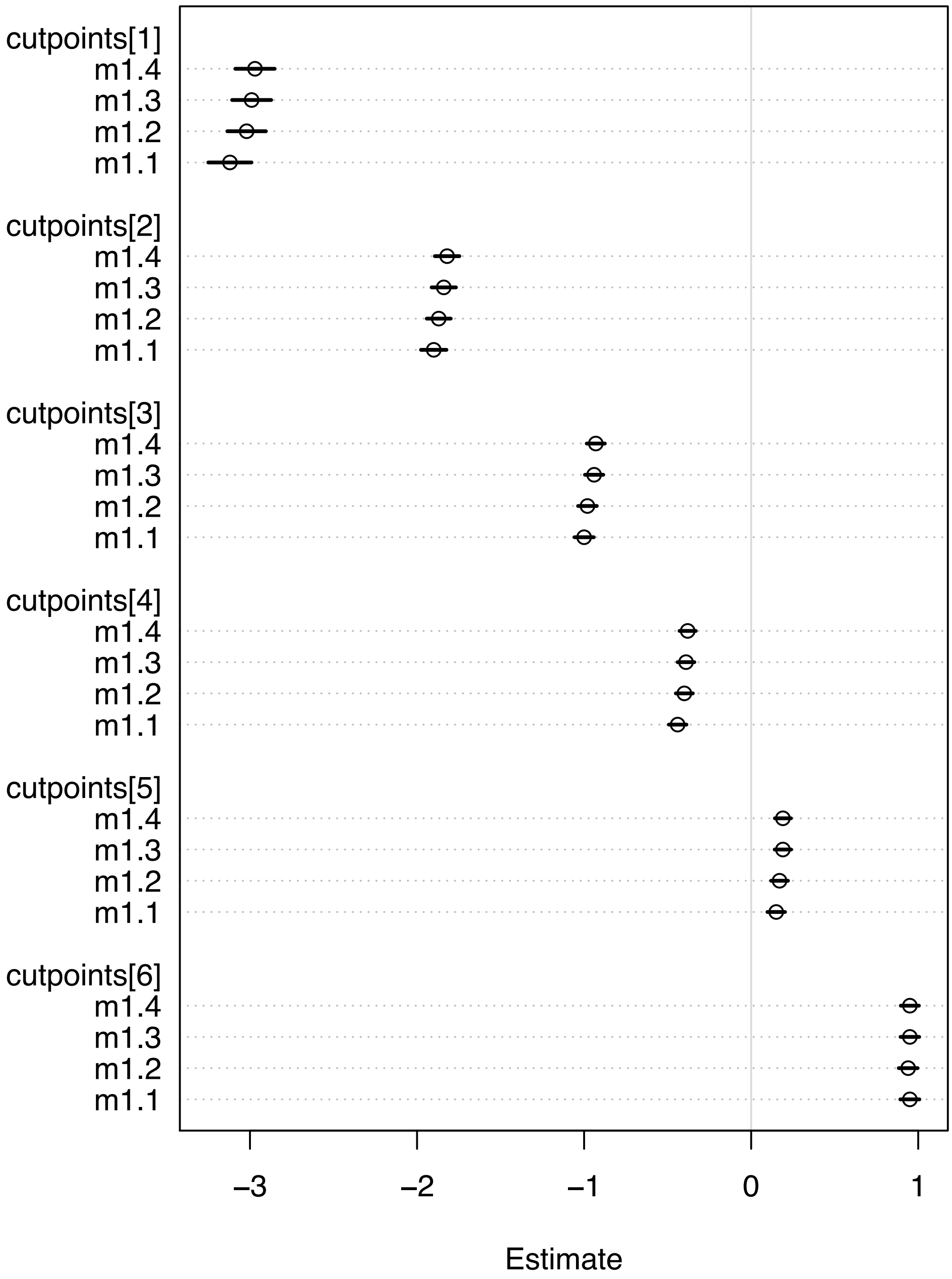


Figure 61: Estimates of the intercepts from Chapter 5 of ordinal categorical model with excluded pairs (without first 10, 20, 30 pairs) that individuals rated.