Connecting through Shared Cognition: Social Consequences and Psychological Underpinnings of Sharing Experiences with Others.

by

Wouter Wolf

Department of Psychology and Neuroscience
Duke University

Date:_______________________

Approved:

Michael Tomasello,
Advisor

Makeba Wilbourn

Paul Seli

Walter Sinnott-Armstrong

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology and Neuroscience in the Graduate School of Duke University

2021
ABSTRACT

Connecting through Shared Cognition: The Social Consequences and Psychological Underpinnings of Sharing Experiences with Others.

by

Wouter Wolf

Department of Psychology and Neuroscience
Duke University

Date:_______________________

Approved:

Michael Tomasello,
Advisor

Makeba Wilbourn

Paul Seli

Walter Sinton-Armstrong

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology and Neuroscience in the Graduate School of Duke University

2021
Abstract

To create social closeness, humans engage in a variety of social activities centered around shared experiences, which, remarkably, do not seem to have a non-human equivalent. Recent work with human adults has suggested that one potential key mechanism through which humans connect to others during shared experiences is shared cognition, the capacity to infer shared mental states, with a particular emphasis on joint attention. To better understand the role of shared experiences, and its underlying social cognition, in human social life, we present a series of studies aimed at examining the ontogeny, phylogeny, contextual flexibility, and (social) consequences of sharing experiences with others through joint attention.

Chapter 1 establishes the importance of the current empirical work by discussing the relationship between shared cognition and social bonding from an evolutionary perspective and describes the current state of the field. It then discusses several fundamental questions that remain unanswered, which form the core of the current dissertation. In Chapters 2-4, we describe a set of studies aimed at better understanding the ontogeny and phylogeny of the role of joint attention in connecting with others through shared experiences. We find that although both human children and great apes behave more socially after co-attending to a video (Chapters 2 and 3), only humans seem to create additional social closeness by creating common ground with their partner
about the fact that they are sharing an experience through mutual gaze in response to a stimulus onset (Chapter 4). In Chapter 5, we describe a study with college students, in which we examine whether the social bonding effect of joint attention also operates in the context of online video mediated interactions, and if this phenomenon is moderated by group size. We find no difference between the joint attention and disjoint attention condition for dyads or groups, suggesting that, regardless of group size, joint attention is not an effective way to create social closeness in video mediated social interactions.

In Chapter 6, we describe a study in which we examine whether children have a social preference for experiencing an activity together, through joint attention, versus alone, and whether sharing this experience shapes children’s attitudes towards the content of that experience in general. Our preliminary data (halted due to Covid 19 safety regulations) show no difference in children’s willingness to stay engaged in a video depending on whether they shared the experience of watching that video or watched the video by themselves (social preference). We did, however, find a trend towards children staying engaged longer during the shared experience. Additionally, we found no effect of joint attention to a toy on children’s motivation to play with that toy during subsequent individual exposure (attitude formation).

In Chapter 7, we examine the darker side of the role of shared cognition in social bonding, namely how, after a shared experience, we are concerned about making a good impression on others. Specifically, we examine the development of the Liking Gap: the
tendency of individuals to, after a brief interaction with a partner, think that their partner evaluated them more negatively than they evaluated their partner. Our results with children between age 4 and 11 show a Liking Gap emerging at age 5, and intensifying between age 5 and 11.

Finally, in Chapter 8 we summarize and synthesize the empirical findings, discuss their theoretical contribution and practical implications, and propose avenues for future research. Overall, these studies demonstrate the crucial role of humans’ sophisticated social cognitive abilities in our social life, enabling us to connect with others effectively through shared experiences. However, our results also suggest that the extent to which this social cognitive mechanism operates outside of the context of face to face interactions might be limited. Finally, the current work highlights that these new opportunities for social bonding might also come with new opportunities to worry about the impression we make on others, even at an early age.
Dedication

For Sylvia,
Who opened the door to science and held my hand until I was old enough to explore it on my own
Ik hoop dat dit werkstuk net zo goed is als dat over Friesland

For Johan,
Whose unwavering support and enthusiasm has carried me through all adversities along the way
Kleine aapjes worden groot

For Merel,
For showing me the way towards the wonderful world of developmental psychology
We lijken misschien toch meer op elkaar dan we denken

For Jens and Joris,
Whose optimism and humor has caused our friendship to grow stronger for over 20 years
Vertel alsjeblieft niemand over de druivensap en de mangosap

And for Jenny,
Who had the courage to marry a scientist and provides the balance to make my life complete
夫妻廢片

vii
# Contents

Abstract ........................................................................................................................................ iv

List of Tables .................................................................................................................................. xiv

List of Figures ................................................................................................................................... xv

Acknowledgements ......................................................................................................................... xvii

Chapter 1. Introduction ...................................................................................................................... 1

The nature of (human) social relationships ...................................................................................... 1

The phylogeny of human social bonding ......................................................................................... 3

Shared cognition. ............................................................................................................................... 6

Shared cognition, shared emotions and social bonding ................................................................. 11

Studying the role of shared cognition in social bonding through joint attention. ...................... 13

Joint attention and social bonding: state of the field and current questions. ............................ 15

Ontogeny .......................................................................................................................................... 17

The ontogeny of joint attention ....................................................................................................... 17

Current research ............................................................................................................................... 20

Phylogeny ......................................................................................................................................... 20

The phylogeny of joint attention ..................................................................................................... 20

Current research ............................................................................................................................... 22

Contextual flexibility ......................................................................................................................... 23

Creating social closeness in online interactions .......................................................................... 23
Current research ................................................................. 23
Group size ............................................................................. 24
The effect of group size on social bonding through shared cognition. .......... 24
Current research .................................................................. 26
Joint attention, social preferences and attitude formation ...................... 26
Social preference and attitude formation in shared social activities .......... 26
Current research .................................................................. 27
The development of the Liking Gap ............................................. 29
The ontogeny of concerns about making a good impression after a shared experience................................................................. 29
Current research .................................................................. 30

Chapter 2. Watching a video together creates social closeness between children and adults ................................................................. 31

Introduction ............................................................................. 31
Study 1 ................................................................................... 35
Methods .................................................................................. 35
Results .................................................................................... 39
Discussion ............................................................................... 43
Study 2 ................................................................................... 45
Methods .................................................................................. 45
Results .................................................................................... 46
Discussion ............................................................................... 48
Chapter 5. Creating social closeness in adult dyads and groups through joint attention in video-mediated social interactions...104

Introduction..............................................................................................................104
Methods ....................................................................................................................109
Participants and design ..........................................................................................109
Procedure ................................................................................................................110
Measures ...............................................................................................................114
Results ....................................................................................................................117
Social bonding........................................................................................................118
Additional control variables ..................................................................................122
Watching experience ...............................................................................................122
Discussion..............................................................................................................123

Chapter 6. Social preferences and attitude formation in children’s shared social activities. ..............................................................132

Introduction............................................................................................................132
Methods ..................................................................................................................136
Participants and Design........................................................................................136
Procedure...............................................................................................................137
Measures and coding .............................................................................................141
Results ....................................................................................................................142
Content attitude ....................................................................................................144
Chapter 7. The development of the liking gap: children over 5 think that partners evaluate them less positively than they evaluate their partners. 154

Introduction 154

Methods 157

Participants and design 157

Procedure 160

Measures 162

Results 164

Discussion 169

Chapter 8: Conclusions 176

Chapter summaries 177

Chapter 2. Co-attending and social closeness in young children 177

Chapter 3. Co-attending and social closeness in great apes 178

Chapter 4. Joint attention, common ground and social closeness in young children and great apes 178

Chapter 5. Connecting over the internet: social bonding through joint attention in online video-mediated dyadic and group interactions 179

Chapter 6. Children’s social preferences for joint attention activities, and the role of joint attention in the formation of attitudes to their content 180

Chapter 7. The development of the Liking Gap 181
Implications and future directions.............................................................182
Joint attention and social bonding in humans ........................................182
Joint attention and social bonding in human children..............................187
Joint attention and social bonding in great apes.....................................196
Mechanisms of shared cognition based social bonding...........................199
Joint attention, social preferences and attitude formation.......................204
The Liking Gap .......................................................................................206
Clinical applications: Shared cognition, social bonding and Autism Spectrum Disorder ...................................................................................208
Coda ........................................................................................................209

Appendix A: Supplementary materials Chapter 3.................................210
Appendix B: Supplementary materials Chapter 4 .................................214
Appendix C: Supplementary materials Chapter 5.................................215
References ..............................................................................................218
Biography ...............................................................................................248
List of Tables

Table 1: Means and standard deviations of latency, looks and bids per condition for the 1.5 year olds in Study 1. ................................................................. 42

Table 2: Means and standard deviations of latency, looks and bids per condition for the 2.5 year olds in Study 1. ................................................................. 42

Table 3: Means and standard deviations of latency, looks and bids per condition for the 2.5 year olds in Study 2. ........................................................................ 48

Table 4: Descriptive statistics of coded behaviors in Study 2. ........................................... 78

Table 5: Model Comparison parameters for social bonding models. .............................. 119

Table 6: Distribution of participants per age group over different types of dyads, and the average age difference per age group. .............................................. 159

Table 7: Model Comparison for models including Age, Age Difference and Dyad Gender Composition. .................................................................................. 165

Table 8: Bayesian parameter estimates for the Liking gap scores at different age groups ............................................................................................................. 167

Table 9: Demographics and number of pairs and trials subjects participated in (Chapter 3 Study 2). ............................................................................................................ 210

Table 10: Great ape demographics (Chapter 4, Study 2) ...................................................... 214

Table 11: Model comparison parameters for impoliteness models. ................................. 215

Table 12: Model comparison parameters for attractiveness models. .............................. 215

Table 13: Model comparison parameters for viewing enjoyment models. ......................... 216

Table 14: Model comparison parameters for video liking models. ................................. 216

Table 15: Model comparison parameters for watching video again models. ................. 217
List of Figures

Figure 1: Distribution of approach latency for 1.5 year olds in the watching together versus control condition (Study 1)..............................................................................................................40

Figure 2: Distribution of approach latency for 2.5 year olds in the watching together versus control condition (study1)..............................................................................................................41

Figure 3: Distribution of approach latency for 2.5 year olds in the watching together versus control condition..........................................................................................................................47

Figure 4: Visual experience of the ape in the Watching Together condition (A) and the control condition (B). Note the eye tracker on the table in front of the television monitor. ...............................................................................................................................62

Figure 5: Approach latency per condition (Study 1).................................................................................................................................66

Figure 6: Set-up of Study 2 for the watching together condition (A) and control condition (B) ..................................................................................................................................................................70

Figure 7: Time spent in the same part of the room (Study 2) .................................................................76

Figure 8: Schematic overview of the manipulation for children (upper box) and great apes (lower box). ................................................................................................................................................................92

Figure 9: Distributions of approach latency per condition for children (left) and great apes (right). ............................................................................................................................................................99

Figure 10: Example of the participants’ perspective during the manipulation for each of the four cells in the design. .........................................................................................................................................................113

Figure 11: Bayesian High Density Intervals containing the 95% most likely social bonding scores for the Joint Attention and Disjoint attention groups per group size.... 120

Figure 12: Set up of manipulation during session 1 (content attitude) ..............................................138

Figure 13: Set up of manipulation during session 2 (social preference).............................................140

Figure 14: Data distributions for Taking leave from the moving toy and Total time spent playing with the moving toy ..........................................................................................................................143
Figure 15: Data distributions for taking leave from video and Total time spent watching the video.................................................................................................................. 143

Figure 16: Bayesian High Density Intervals containing the 95% most likely values of the condition difference scores for first leave and total time played.............................................. 145

Figure 17: Bayesian High Density Intervals containing the 95% most likely values of the condition difference scores for first leave and total time played.............................................. 147

Figure 18: Smiley face Likert Scale. ................................................................................................................................. 162

Figure 19: Bayesian 95% HDI for Liking Gap scores at different age groups. .................... 168

Figure 20: Individual distribution of approach latency, seperated by species (Chapter 3, Study 1).................................................................................................................................................................................. 211

Figure 21: layout of the rooms (Chapter 3, Study 2)................................................................. 212

Figure 22: Room division during coding of Same part of the room measure (Chapter 3, Study 2).................................................................................................................................................................................. 213
Acknowledgements

Conducting research with data from three types of populations (human children, adults and great apes) collected over three continents (US, Europe and Africa) was a daunting enterprise, and would have been an impossible had it not been for the incredible support I received over the last few years.

First and foremost, I would like to thank my advisor and mentor, Professor Michael Tomasello. Without his guidance and support this work would not have been possible. From our very first meeting I felt an instant connection through shared interests and a similar way of thinking. I will always remember and cherish our discussions about the social cognitive development of young children and great apes. Thank you for your help and advice, and for trusting me and believing in me. Despite our lab only consisting of three members at the very beginning, the two of us and Paige, our lab manager, you made sure I always felt at home.

I would also like to thank my committee members, Dr. Makeba Parramore Wilbourn, Dr. Walter Sinnott-Armstrong, and Dr. Paul Seli, for dedicating their time to provide feedback and support me during various stages of the project, as well as being willing to work with me through the impact of the Covid 19 pandemic. I would also like to thank Dr. Makeba Parramore Wilbourn and Dr. Steve Asher for being a mentor in
teaching and academia as a whole. Our conversations have caused me to look at myself differently, both as an academic and as a person.

The current research would not have been possible without the financial support from Duke’s graduate school, their Lafitte research and travel grants, and the Philip Jackson Baugh Fellowship for Research in Human Development I was awarded in my final year. In addition, I would like to thank the Fulbright Foundation and Institute of International Education, whose scholarship, as well as guidance in navigating the perils of being an international student have proven invaluable.

Furthermore, I would like to thank all my colleagues and friends who have provided me with valuable feedback about my ideas and research. In particular, I would like to thank my two fellow PhD students, Jared Vasil and Leon Li, for their feedback on my studies, Christopher Krupenye, for his feedback on my ape studies and the use of eye tracking in these studies, Robert Hepach, for helping me analyzing my eye tracking data, Jan Engelmann, who provided valuable feedback on my initial research designs, Juliane Kaminski, without whom I do not know how I would have completed my data collection in Uganda, Brian Hare, who welcomed me into his lab meetings to discuss my research, which was always a delightful experience, and all the other researchers and students at Duke University, in the developmental area and beyond, as well as those at the Max Planck Institute for Evolutionary Anthropology in Leipzig for their feedback on my ideas.
Finally, this project would not have been possible without the people from the Tomasello Lab, including our lab managers, Paige, Amanda and Julia, my honors thesis students, Rachel, Julia and Kayley, and all the research assistants who helped me run studies over the last couple of years. In addition, I would like to thank all the supporting staff at Duke, as well as the staff and research assistants at the Max Planck Institute for Evolutionary Anthropology in Leipzig and the Leipzig Zoo, and the Ngamba Island Chimpanzee Sanctuary, for their help in data collection. And, of course, I would like to thank all the children, parents, students and apes whose participation was instrumental in the completion of this dissertation research.
Chapter 1. Introduction

Creating and maintaining social bonds is a crucial part of human social life, and failing to do so leads to reduced mental and physical well-being (Baumeister & Leary, 1995). Humans are not alone in that respect. For many social animals, including non-human primates, being ostracized or failing to form an alliance within a group is a major source of stress (Sapolsky, 1994). All social primates therefore spend a considerable amount of time on social bonding behavior, taking away time from other essential activities such as feeding (Dunbar, 1991; Lehmann et al., 2007). In fact, social bonding is so costly in terms of time that research on primate time budgets has suggested that time is the main factor constraining the number of social relationships that can be maintained within a species (Dunbar, 1991; Dunbar et al., 2009; Lehmann et al., 2007). Yet, although there are only 24 hours in a day for all primates, humans manage to form and maintain considerably larger social networks than other primates (Dunbar, 2020). This raises the question how humans are capable of forming and maintaining such large social networks.

The nature of (human) social relationships

To understand how humans form and maintain such a relatively high number of social relationships, it is important to clarify what we mean by a social relationship. To many, the idea that humans have relatively large social networks might seem counterintuitive. After all, there are many species that live together in much larger
groups than humans, such as flocks of birds and schools of fish. Yet, many researchers have argued that a social network entails more than just a large number of aggregated individuals living together (Shultz & Dunbar, 2007). That is, one important feature of social networks is stability (Shultz & Dunbar, 2007). For example, flocks of birds and schools of fish are highly fluid in terms of membership, with individuals joining and leaving in accordance with their own personal needs and/or environmental demands. Yet, in for example primate groups, membership is relatively stable. In such groups individuals interact with each other on a regular basis, creating idiosyncratic, stable patterns of social behavior and social expectations (i.e., social relationships) between individual group members (Dunbar, 2020).

The evolutionary history of the emergence of stable group membership through social bonds, at least in primates, has been proposed to have occurred in several steps (Shultz et al., 2011): From solitary foraging individuals to pair-bonded dyads, followed by social aggregations of large multi-male/multi-female groups which included pairbonding or single male harem systems, and then finally a shift to stable socially bonded groups that include alliances and friendship networks (Silk, 2002, 2007). In this latter stage in particular, maintaining stability in a social network requires active social bonding, meaning that individuals have to reserve time for social interactions aimed at maintaining their social relationships (Shultz & Dunbar, 2007, 2010). As this is also the stage in which humans seemed to have set themselves apart from other animals, the
answer to the question why humans are capable of maintaining considerably larger social networks than our closest primate relatives (Dunbar, 1998, 2008, 2012, 2020; Lehmann et al., 2007) might therefore be found in the unique ways in which humans create and maintain social relationships with others.

**The phylogeny of human social bonding**

Humans have many social bonding mechanisms in common with other animals. On a very basic level, in many mammals, social network structure is associated with physical proximity (Ancillotto et al., 2015; Arseneau-Robar et al., 2018; Micheletta et al., 2012; Schamberg et al., 2016; Zhang et al., 2012). Although it has never been experimentally tested whether physical proximity without further interaction creates social closeness between individuals, it seems plausible that familiarity through mere exposure causes animals to create more favorable attitudes towards others the same way it does in humans (Bornstein, 1989; W. Wilson & Miller, 1961; W. Wilson & Nakajo, 1965; Zajonc, 1968).

Another mechanism of generating social closeness in humans as well as non-human primates is behavioral mimicry: the tendency to adopt the postures, gestures and mannerisms of those we interact with (Chartrand & Bargh, 1999). It has been well documented that, in humans, mimicry increases feelings of affiliation and rapport (Chartrand & Bargh, 1999; Chartrand & Lakin, 2013; Chartrand & van Baaren, 2009; Kühn et al., 2010). More recently, studies have shown that mimicry has a similar effect in
monkeys (Paukner et al., 2009). However, whether this ‘chameleon effect’ has an actual impact on non-human primates’ social life is still debated, as spontaneous imitation seems to occur less often in non-human primates than in humans (Call & Carpenter, 2003, 2009; Paukner et al., 2009).

Although proximity and mimicry implicitly facilitate social closeness, many species, in addition, create social closeness by engaging in social activities in which physical touch plays a crucial role. Harlow’s classic studies (Harlow, 1958) already showed that even young rhesus macaques were motivated to spend the majority of their time close to maternal models that provided a pleasant tactile experience (over a model that provided food), a response, he suggested, originating from a need for ‘love’ (i.e., attachment). Subsequent research corroborated the idea that physical contact is a crucial part of primate social bonding, in particular through grooming, non-procreative sex and physical play (Arseneau-Robar et al., 2018; Behncke, 2015; Dunbar, 1991; Hare et al., 2015; Lehmann et al., 2007; Wittig et al., 2008). A similar role of touch in creating social closeness has been found in humans (Dunbar, 2010; Gallace & Spence, 2010, 2016). Moreover, studies have shown that the appropriateness of being touched by others is moderated by one’s social relationship to them (Suvilehto et al., 2015), further demonstrating the relationship between touch and social closeness in humans.

Yet although we share these previous social bonding mechanisms with our closest living evolutionary relatives, humans have certain social bonding activities for
which no non-human equivalents seem to exist. Most notably, we connect with others through language. Research has shown that disclosing personal information with others is an effective way of generating social closeness (Aron et al., 1997). But the information being shared does not need to be self-related in order for social bonding to occur. That is, exchanging social information about any individual within a social network, often referred to as gossip, has been shown to be an equally effective social bonding mechanism (Bosson et al., 2006; Dunbar, 2004). In addition, humans also engage in many social bonding activities that are not centered around conversations. Research has shown that making music together creates social closeness (Pearce et al., 2015; Weinstein et al., 2016), as well as playing team sports (Artinger et al., 2006), playing cooperative (video) games (Dabbish, 2008; Depping & Mandryk, 2017; Zheng et al., 2002), dancing together in synchrony (Cirelli et al., 2017; Hove & Risen, 2009; Tarr et al., 2015, 2016) and participating in rituals (Charles et al., 2021; Singh et al., 2020).

Crucially, these social bonding activities are all centered around creating a shared experience with others for social bonding purposes, and do not seem to have a non-human equivalent. It has therefore been argued that such activities have played a crucial role in the evolution of human sociality: not only is creating social closeness through shared experiences safer than through physical interaction (e.g., reducing the risk of pathogen transmission), but also more efficient, as it can be done with multiple individuals at the same time (Dunbar, 1998, 2004, 2008, 2012). In this discussion on the
evolutionary origin of human social bonding, the exact nature of the psychological underpinnings that allow us to connect with others through such activities remained mostly unexplored until recently researchers started pointing out the potential key role of humans’ shared cognition in social bonding (Wolf, Launay, et al., 2015).

**Shared cognition.**

Over the last decades, scholars from different disciplines have aimed to better understand the psychology underlying humans’ uniquely effective cooperation and communication. In doing so, many have pointed out that humans’ behavior coordination is undergirded by a relatively sophisticated cognitive capacity to understand and coordinate mental states, as well as a strong motivation to do so. Together this cluster of cognitive mechanisms and motivations related to coordinating mental states is often referred to as shared intentionality, while the cognitive aspect of this construct is sometimes referred to as shared cognition (Call, 2009; Rakoczy & Tomasello, 2007; Tomasello, 2014, 2019; Tomasello & Carpenter, 2007; Tomasello & Rakoczy, 2003). Yet, despite (or perhaps due to) the long history and broad interest in the role of shared cognition in human coordination and communication, there is no clear consensus on a mechanistic definition of the term. We will therefore briefly conceptualize the meaning of shared cognition in the context of the current work.

In line with previous work, we use the term shared cognition to describe the cognitive capacity of individuals to make specific inferences about others’ mental states
in a way that allows them to better understand and predict the behavior of those around them (Shteynberg, 2015, 2018; Shteynberg et al., 2020; Tomasello, 2014, 2019; Tomasello & Carpenter, 2007; Tomasello & Rakoczy, 2003). Additionally, the term shared experience will from here on be used to refer to social experiences during which individuals use this shared cognition while interacting with others. Acquiring this capacity provides individuals with a substantially different understanding of their social environment, which has a profound influence on individual behavior as well as on interpersonal and group dynamics. However, to understand the role of shared cognition in social bonding, we must first understand how individuals’ shared cognition facilitates their individual feelings of closeness towards others, meaning that the primary level of analysis in this work is that of the individual.

Crucially, then, in the current work we construe shared cognition as the capacity to make four types of inferences with regard to the mental states of others: general Theory of Mind inferences, inferences of mental state similarity, inferences of mental state convergence and common ground inferences. First, shared cognition requires an individual to make inferences about the mental states of others. In other words, shared cognition presupposes Theory of Mind (Herzog et al., 2013). Second, shared cognition requires one to compare these mental states of others to one’s own, inferring the degree of mental state similarity between oneself and a partner. Third, even if we infer that the mental states we are experiencing are similar to that of a partner, we also need to infer
that these mental states are directed towards the same information. After all, two people can experience a similar level of sadness about completely different things, or be similarly angry with one another, in which case they will not infer mental state convergence.

Fourth, one must be able to infer common ground. Common ground generally refers to a set of inferences and assumptions about an interaction partner’s knowledge about information relevant to that interaction, and the mutual availability of that knowledge (Bohn & Köymen, 2018; Clark et al., 1983; Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986; Holler & Wilkin, 2009; Lee, 2001; Liebal et al., 2013; Siposova & Carpenter, 2019; Tomasello, 2014). A common way to describe how this mutuality of knowledge in common ground is inferred is through higher order recursive mindreading (Oesch & Dunbar, 2017; O’Grady et al., 2015; Tomasello, 2014, 2019), reminiscent of Grice’s recursive account on inferences of meaning in language (Grice, 1957). In line with this account, the current work conceptualizes common ground inferences in the context of shared cognition as an individual inferring that their partner has inferred that both they and their partner are aware of them experiencing a similar mental state towards the same information.

With regard to these four types of inferences, two issues need to be briefly addressed. First, when it comes to inferences of mental state similarity and convergence the problem is that the mental states of two different individuals can never be truly
identical. Even if two people are looking at the same object, they will be looking at it from different perspectives (meaning that their perceptions will slightly differ) and even if they are both happy about seeing that object, there is no way of telling whether they are ‘equally’ happy. This level of accuracy is, however, also not necessary for these inferences to benefit an interaction. Instead, a more functional approach to the level of accuracy of these inferences, in line with theoretical work on grounding and relevance in communication (Clark & Brennan, 1991; D. Wilson & Sperber, 2002), is that individuals will deem their mental states and those of their interaction partner sufficiently similar and convergent if that level of similarity and convergence is sufficient for the purposes of that interaction. A similar type of reasoning can be applied to common ground inferences: one only needs to be sufficiently certain of the mutual awareness about a mental state being shared with another individual to act on this common ground when interacting with this individual.

Second, an important issue that needs to be addressed here concerns the way in which shared cognition inferences are processed, which has been subject of much debate. The importance of this issue is clearly illustrated by developmental work examining when Theory of Mind emerges in childhood. For a long time researchers have argued that Theory of Mind emerges in childhood between 3 and 4 (Herzog et al., 2013; Perner et al., 2002; Wellman et al., 2001, 2001; Wimmer & Perner, 1983), with children consistently passing false belief tasks at age 4. Then, new evidence appeared
suggesting that Theory of Mind already emerges in early infancy (for reviews, see: Baillargeon et al., 2010, 2016; Scott & Baillargeon, 2017). Although some of this work suffered from methodological confounds, causing results not to replicate in subsequent studies (Dörrenberg et al., 2018; Powell et al., 2018), other tasks do seem to reliably show Theory of Mind emerging at 2 years of age, and perhaps even earlier (Powell et al., 2018).

The crucial difference between these more recent studies on Theory of Mind and their earlier counterparts is that the more recent studies attempt to reduce the degree to which the capacity to represent mental states at a conceptual level is required to successfully complete the task (Powell et al., 2018). Importantly, this line of reasoning thus presupposes that Theory of Mind does not necessarily require mental states to be represented on a conceptual level. Instead, it is argued that Theory of Mind inferences can be made through more automated heuristic or schema based processing of subdoxastic informational states about the mental states of others (rather than conceptually informed beliefs). This would allow younger children to adapt their behavior to the mental states of others without them being able to conceptually represent these mental states (Rakoczy, 2012).

With regard to the current work, it is therefore important to note that this idea can be applied to all shared cognition inferences. That is, shared cognition inferences can be made by processing mental states, their convergence and similarity, and common
ground awareness subdoxastically, without being able to represent any of these elements, or their recursive relationships, at a conceptual level. This might explain why children have been shown to act in congruence with common ground (i.e., assumptions about mental states related to each other in a recursively structured way) established with a partner as early as 14 to 18 months (Liebal et al., 2009, 2010, 2013; Moll et al., 2007, 2008), while they struggle to successfully solve more conceptually heavy second order Theory of Mind tasks until age 6 (Grueneisen et al., 2015; Perner & Wimmer, 1985) and third order Theory of Mind tasks until age 10 (Liddle & Nettle, 2006).

**Shared cognition, shared emotions and social bonding**

Over the last couple of decades, a rich literature emerged demonstrating that shared cognition inferences allow humans to more accurately predict and anticipate the behavior of others, better coordinate their mental states by communicating about them and, in doing so, effectively assimilate and transmit culture (Angus & Newton, 2015; Herrmann et al., 2007; Kern & Moll, 2017; Moll & Tomasello, 2007; Tomasello, Carpenter, Call, et al., 2005; Tomasello, 2010b, 2010a, 2014, 2019; Tomasello & Carpenter, 2007; Tomasello & Herrmann, 2010; Tomasello & Vaish, 2013). As such, shared cognition can be considered one of the cornerstones of human sociality. However, the idea that this foundational cognitive capacity might also shape human social bonding remained unexplored for a long time, in part because most social bonding research mainly focused
on identifying different behaviors that facilitate social bonding, rather than exploring the social cognitive mechanisms underlying the social bonding effect of such behaviors.

Some scholars have, however, explored the role of synchronizing affect during social bonding activities, in particular through social laughter. This research suggests that social laughter might facilitate social bonding through the release of opioids in the brain (Caruana, 2017; Dezecache & Dunbar, 2012; Dunbar et al., 2012; Fraley & Aron, 2004; Manninen et al., 2017). Most of these studies, however, lack control conditions in which individuals laugh individually, making it impossible to determine whether it is the sharedness of the affect, or the affect itself that causes social bonding in these situations. After all, it is not implausible that laughter puts participants in a better mood, which, in turn, might cause them to form more favorable attitudes towards their partner.

Some researchers have tried to address this issue by including additional control conditions in their design (Bazzini et al., 2007; Kurtz & Algoe, 2017). However, in some cases these studies had control conditions that still included some degree of shared laughter (Kurtz & Algoe, 2017, Study 3), or had participants remembering (non-) shared laughter (or other shared positive events) together, creating a novel shared experience in the process (Bazzini et al., 2007). As such, these control conditions still caused participants to make shared cognition inferences, making it difficult to isolate the effect of making such inferences on social closeness.
Furthermore, most of these studies did not measure social bonding or social closeness, but, instead relationship satisfaction (i.e., in the context of previously established social relationships). In one study, for example, participants were asked how satisfied they were with a relationship after recalling episodes of (non-) shared laughter with that person (Kurtz & Algoe, 2017, Study 2). The problem here is that asking participants about relationship satisfaction without asking them to specify the nature of that relationship makes it difficult to draw conclusions about social closeness. After all, one can be very satisfied with having either a close or distal social relationship with a specific person. Furthermore, the Bazzini et al. (2007) study was aimed at examining the effect of shared laughter on satisfaction in romantic relationships. They therefore measured this satisfaction not only by asking participants directly about this relationship satisfaction, but also by asking them about their partner’s attractiveness and the degree to which they were ‘in love’ with their partner, both of which seem to be closer related to romantic affiliation motivation than social bonding in general. Thus, although the literature on shared laughter suggests that these experiences have important social consequences, it is difficult to draw conclusions from this literature about the role of shared cognition in social bonding specifically.

**Studying the role of shared cognition in social bonding through joint attention.**

To assess the role of shared cognition in social bonding more directly, researchers recently set out to examine the effect of joint attention on subsequently experienced
social closeness (Wolf, Launay, et al., 2015). One important reason to study shared cognition in the context of joint attention is that controlling participants’ attention experimentally is relatively easy. To do so, one can simply manipulate the visual field of participants and/or the stimuli presented to them. Furthermore, manipulating the inferences participants make about others’ attention is also relatively straightforward, for example by visibly manipulating the visual field of this partner. Another possibility is having confederates avert their gaze from a stimulus. This latter approach should be effective even when conducting research with young children and great apes, as research has shown that both of these populations can infer what a partner can and cannot see from the direction of a this partner’s gaze (Bräuer et al., 2007; Grossmann & Johnson, 2010; Hare et al., 2000; Karg et al., 2015; Moll & Tomasello, 2004; Nuku & Bekkering, 2008; Tomasello, Hare, et al., 2007).

Studying the effect of joint attention on social bonding is, however, not only convenient, but also of great theoretical importance: the capacity to engage in joint attention is a pivotal step in the development of shared cognition and shared intentionality, and is crucial in the acquisition of the skills necessary for cooperation and communication (Baldwin, 1995; Bruner, 1995; Butterworth, 1995; Moore & Dunham, 1995; Mundy, 2006; Mundy & Newell, 2007; Scaife & Bruner, 1975; Tomasello, 1995, 2010a, 2014, 2019). In fact, one of the earliest signs of developmental disorders related to communicative and cooperative cognition, such as Autism Spectrum Disorder, is the
delay (or absence) in the development of joint attentional skills (Charman, 2003; Charman et al., 1997; Dawson et al., 2004; Gomez, 2005; Jones & Carr, 2004; Mundy, 2006; Mundy et al., 1986, 1992, 1994; Mundy & Newell, 2007). As such, studying the role of joint attention in creating social closeness does not only have theoretical implications for how shared cognition has shaped human social bonding, but might also have important clinical applications.

**Joint attention and social bonding: state of the field and current questions.**

In a recent experiment with adults, researchers first showed that joint attention elicits social closeness in adults (Wolf, Launay, et al., 2015). Participants came into the lab in a group of four after which they were seated behind individual computers. Here, they received instructions and training in a response time task. They then engaged in two sessions of this response time task, each time with a different participant sitting next to them. Crucially, in one session participants attended to the same stimulus as their partner, whereas in the other session, participants responded to stimuli on opposite sides of the screen. After each session, participants were brought back to their individual computers and answered a series of questions, including questions about how they felt about their partner. Results showed that participants felt closer to partners with whom they had been attending to the same stimuli than towards partners who were consistently looking at a different part of the screen. Subsequent research replicated this effect in a video watching context, while also showing that this effect is stronger when
stimuli elicit higher arousal (Dunbar et al., 2016; Rennung & Göritz, 2015) and disappears when the stimulus contains information that threatens one’s ingroup (Haj-Mohamadi et al., 2018).

These studies provide an important first indication that joint attention, and thereby shared cognition, facilitates the creation of social closeness between humans. Yet, to truly understand how foundational this social bonding mechanism is to our species’ capacity to form and maintain social networks, and gain more insight into its psychological underpinnings and social consequences, six pivotal questions still require answering:

1. Does the social bonding effect of joint attention already emerge in childhood? (ontogeny)

2. Is the social bonding effect of joint attention uniquely human, or can (a form of) of this phenomenon also be found in other species? (phylogeny)

3. How flexibly does our shared cognition create social closeness in different contexts? Does joint attention still elicit social closeness in interactions without physical proximity, such as online video-mediated interactions? Or is the social bonding capacity of joint attention constrained to face to face interactions? (contextual flexibility)

4. Does shared cognition actually allow us to connect effectively with multiple individuals at the same time? Do we connect with other individuals to the same
degree when we engage in joint attention in a dyadic interaction compared to when we engage in joint attention with a larger group? (group size)

5. Do children prefer joint attention activities over individual activities? And is this preference constrained to the social context of that activity? Or does joint attention also shape children’s attitudes towards the content of an activity even to the extent that they are more motivated to engage in a similar activity by themselves in the future? (social preferences and attitude formation)

6. How do the negative psychological consequences of engaging in a shared experience develop? Research has shown that after engaging in a shared experience with a partner, people evaluate their partner more positively than they think their partner evaluates them, a phenomenon referred to as the Liking Gap. When does this Liking Gap emerge in ontogeny? And how does it develop over childhood? (Liking Gap)

**Ontogeny**

**The ontogeny of joint attention**

If young children already connect with others through joint attention, they have to be able to (1) infer the attention of others, (2) infer that both they and their partner are being attentive, (3) infer that their attention has converged with that of a partner, and (4) establish common ground about the fact that they and their partner are both attending to the same thing. Children develop the capacity to infer others’ attention and converge
their attention with that of their partner early in ontogeny. Already at 2 months of age children start adjusting their gaze based on a change of focus in adults’ attention (Scaife & Bruner, 1975). At 6 months, infants will follow the general direction of others’ gaze, but will stop when they have reached the first object in the line of this gaze, and will only do so when there is an object towards which the gaze can be directed in their initial visual field. This means that they for example do not follow gaze towards things behind them (Butterworth & Cochran, 1980). Infants’ gaze tracking abilities continue to improve as they get older, where they localize things correctly more often at 12 months (but still struggle in large spaces; Butterworth & Jarrett, 1991), while at 18 months infants reliably follow gaze to space behind them if there is nothing interesting in front of them (Butterworth, 1991, 1995).

More recently, research has shown that even at 12 months children can follow gaze to locations behind opaque barriers, not only suggesting that they can follow gaze to locations outside of their initial field of sight, but also that they seem to have some understanding that the looking behaviors of others are acts of seeing (Moll & Tomasello, 2004). Finally, between 6 and 18 months, the amount of naturally occurring joint engagement between infants and their parents, in which infants consistently keep track of their partner’s attention as well as the object they are both attending to, increases steadily (Adamson & Bakeman, 1985; Bakeman & Adamson, 1984).
From 12 months on, children also seem to make some form of common ground inferences, demonstrated by the fact that they understand that others are intentionally communicating their attention to them, and that they can do this for others as well. Children start to use pointing in order to redirect their partner’s attention at around 12 months (Behne et al., 2012; Chen et al., 2009; Cochet & Vauclair, 2010; Liszkowski et al., 2004, 2007a, 2007b; Murphy, 1978; Tomasello, Carpenter, et al., 2007, 2007). Furthermore, research has shown that around the same time, children also start responding to points. Researchers found that, at around 12 months, most infants reliably fixate on targets their caregivers point to (Butterworth, 1995), although some have argued that at 12 months infants look more at the general direction the adult points to, whereas at 15 months they look more precisely at the object the point was intended to refer to (Morissette et al., 1995).

Crucially, from 14 months on, children seem to be able to process the communicative intent of points. That is, at this age, studies show that a point by an experimenter preceded by a look at the child and then complemented with a look to an object causes children to grab that object more often than when the experimenter’s pointing gesture was not preceded by a look to the child and complemented with a look to a watch on the experimenter’s arm (Behne, Carpenter, & Tomasello, 2005). In addition, research has shown that at this age infants take into account previous experiences with the experimenter to decide what their point actually refers to (Behne et
al., 2005; Liebal et al., 2009; Moll et al., 2007, 2008), making an even stronger case for infants at that age making common ground inferences.

**Current research**

As such, it seems that by the time children are 18 months old, they are capable of reliably making joint attention inferences. Thus, in order to examine whether young children create social closeness through joint attention, we conducted a series of studies, using methodology similar to the original adult study (Wolf, Launay, et al., 2015). In this study, 18 month (1.5 year) olds and 30 month (2.5 year) olds watched a video while sitting next to an experimenter who was either watching with them or reading a book instead, after which we measured social closeness. More specifically, we conducted one study in the context of a natural joint attention interaction (Chapter 2, Study 1), one study designed to isolate the effect of the convergence inference (Chapter 2, Study 2), and one study designed to isolate the effect of the common ground inference (Chapter 4, Study 1).

**Phylogeny**

**The phylogeny of joint attention**

Generating hypotheses about whether non-human animals might be able to connect through joint attention is not an easy task. The capacity of non-human animals to make Theory of Mind inferences alone has already been a subject of intense debate for multiple decades (Call & Tomasello, 2008; Heyes, 1998; Povinelli & Vonk, 2003; Premack
& Woodruff, 1978). In recent years, however, empirical work has provided converging evidence that non-human animals have Theory of Mind capabilities (Amici et al., 2009; Bugnyar & Heinrich, 2006; Buttelmann et al., 2007; Call et al., 2004; Call & Tomasello, 1998; Canteloup & Meunier, 2017; Dally et al., 2004, 2006; Emery & Clayton, 2001; Flombaum & Santos, 2005; Hare et al., 2000, 2001; Kaminski et al., 2005, 2013; Kano et al., 2019; Karg et al., 2015; Krupenye et al., 2016; Kundey et al., 2010; Phillips et al., 2009; Santos et al., 2006; Scerif et al., 2004; Schmelz et al., 2011; Tomasello, Carpenter, & Hobson, 2005; Tomasello, Hare, et al., 2007; Udell et al., 2011). Specifically, in relation to joint attention, non-human animals are able to reliably infer what others can and cannot see (Bräuer et al., 2007; Bugnyar & Heinrich, 2006; Dally et al., 2004, 2006; Emery & Clayton, 2001; Hare et al., 2000; Kaminski et al., 2005, 2013; Karg et al., 2015; Udell et al., 2011).

Non-human animals, including great apes, also seem to easily converge their attention with others. (Amici et al., 2009; Tomasello et al., 1998, 1999; Tomasello, Hare, et al., 2007), although, in doing so they seem to rely more on the orientation of their partner’s head than the movement of their eyes (Tomasello, Hare, et al., 2007). Whether non-human animals can make common ground inferences is, however, not as clear cut. For example, great apes seem to struggle to interpret pointing gestures to locate food (Herrmann & Tomasello, 2006; Tempelmann et al., 2013; Tomasello et al., 1997), although others have found that they can do so successfully (Call & Tomasello, 1994;
Mulcahy & Call, 2009). Furthermore, there is much evidence that non-human animals produce point-like gestures or other intentional signals to shift a partner’s attention to request food (Anderson et al., 2010; Cartmill & Byrne, 2010; Gaunet & Deputte, 2011; Halina et al., 2013; Hattori et al., 2010; Hobaiter & Byrne, 2011a, 2011b, 2014; Kaminski et al., 2011; Leavens et al., 1999a; Liebal et al., 2004; Savalli et al., 2014; Tomasello, 2010a, 2019). The extent to which they understand the pointer’s communicative intent, or are able to infer that individuals reacting to their gestures have inferred their intention to inform is, however, still heavily debated (Tomasello, 2010a, 2014, 2019; Tomasello & Carpenter, 2007).

**Current research**

Overall, the literature provides an ambivalent picture with regard to the degree to which non-human animals, including great apes, are capable to make the inferences required to engage in joint attention. Thus, in our attempt to assess the degree to which non-human animals create social closeness through joint attention, we conducted three studies with great apes, two of which were designed to isolate the effect of attention convergence on social bonding (Chapter 2, Study 1 and 2) and one of which was designed to isolate the effect of common ground inferences on subsequent social closeness (Chapter 4, Study 2).
Contextual flexibility

Creating social closeness in online interactions

If shared cognition facilitates social bonding, then every novel technology aimed at facilitating human cooperation and communication introduces novel opportunities to create social closeness with others. The question is, however, whether our social cognition is flexible enough to keep up with these new modes of social interaction. For example, given that physical proximity has played an important role in social bonding throughout our evolutionary history, it might be that removing this physical proximity from social interactions, as for example happens in online social interactions, might make social bonding more difficult.

Research has shown, however, that even minimalistic social signals in an online environment, such as social media ‘Likes’ can influence participants sense of belonging (Wolf, Levordashka, et al., 2015). Furthermore, online social interactions have been shown to facilitate trust (Zheng et al., 2002) and increases trust and positive attitudes between individuals (Dabbish, 2008; Depping & Mandryk, 2017) while playing games with others online.

Current research

Although the literature suggests that our shared cognition can still influence partner evaluations through online social interactions, through cooperation and feelings of interdependence, the question remains whether creating this social closeness through
joint attention alone is still effective in online social interactions. We therefore set out to explore the extent to which joint attention in an online video-mediated social interaction creates social closeness between college students. Participants either jointly attended to a video on a shared screen with (pre-recorded) confederates, or were told that they would all watch the video sequentially, and watched it by themselves while the other participants were attending to their phones (Chapter 5). We then measured participants’ perceived social closeness through a survey.

**Group size**

**The effect of group size on social bonding through shared cognition.**

As stated earlier, the capacity to connect through shared social activities has been argued to be a game changer for human social bonding and our social network size because these activities allow us to create social closeness with multiple individuals at the same time (Dunbar, 1998, 2004, 2008, 2012). Remarkably, however, there is only one study that directly examines the effect of group size on social bonding in social activities, specifically in the case of singing (Weinstein et al., 2016). This study showed that, when comparing the social closeness of individuals in small (20-80 individuals) and large (232 individuals) musical groups before and after a singing session, the change in social bonding pre to post singing was greater in the larger group than in the smaller group. Importantly, however, this effect was most likely confounded by the fact that the small
singing groups were pre-established social networks, while the large singing group (for the most part) was not.

Due to the pre-existence of social networks in these samples, it is difficult to draw strong conclusions about the effect of group size on social bonding in shared experiences. Furthermore, even if this research was done with strangers, the effects might be specific to making music together. Additionally, comparing smaller groups to larger groups in general does not directly address the evolutionary argument that human social activities are more efficient social bonding mechanisms than those employed by other animals. After all, most of the core social bonding mechanisms of other primates, such as grooming and non-procreative sex are dyadic in nature. It is therefore important to compare the social closeness elicited by joint attention in dyadic shared experiences to the social closeness elicited between individuals during experiences in which multiple members in a group engage in joint attention. Unfortunately, although there are both studies that have investigated whether joint attention elicits social closeness in dyads (Haj-Mohamadi et al., 2018; Wolf, Launay, et al., 2015) and in a small group (Rennung & Göritz, 2015), none of these studies manipulated group size directly within their design, and the measures used in these studies were too different to compare effects across studies.
Current research

We therefore set out to explore the effect of group size on the social bonding effect of joint attention. To do so we added two additional conditions to the study on joint attention and social bonding in video-mediated social interactions described in the previous section (Chapter 5). Participants in this study engaged in the experiment with either one or three (pre-recorded) confederates. This allowed us to directly compare the social bonding score of participants who interacted with one participant (i.e., in a dyad) to average social scores of participants who interacted with three participants (i.e., in a group), and examine the total social closeness generated during these interactions.

Joint attention, social preferences and attitude formation

Social preference and attitude formation in shared social activities

Humans have a fundamental need to belong (Baumeister & Leary, 1995), which seemingly manifests itself in a preference to collaborate. Studies have shown that young children, but not great apes prefer to collaborate as opposed to doing a task individually (Bullinger et al., 2011; Rekers et al., 2011). Yet, does this preference stem from the interdependence and cooperation inherent to successfully completing a collaborative task? Or could simply sharing an experience in joint attention in itself make the activity more enjoyable? And if sharing an activity makes that activity more enjoyable, is this enjoyment constrained to the activity’s social context? Or does sharing an activity in
joint attention shape one’s general attitude to the content of that activity to the point that one is motivated to engage in a similar activity again in the future, even by oneself?

Research on how joint attention shapes perceptions and attitudes towards stimuli has shown that jointly attending to stimuli not only makes those stimuli easier to remember (Eskenazi et al., 2013), but also causes them to be perceived as more extreme regardless of their valence (Boothby et al., 2014, 2017; Shteynberg, Hirsh, Apfelbaum, et al., 2014; Shteynberg, Hirsh, Galinsky, et al., 2014). Yet, in these studies participants’ attitude towards the object and their enjoyment of the social aspect of the interaction were not measured separately.

In addition, the question whether children’s attitudes towards stimuli and/or activities are shaped by shared cognition has remained mostly unexplored. There is, however, some evidence that watching something together as opposed to alone at least moderates children’s arousal and readiness for information processing: Two studies showed that watching a video together with a parent (as opposed to without a parent) was associated with higher skin conductance levels and lower heart rate in children between 6 and 13 years old (Keene et al., 2019; Rasmussen et al., 2017).

**Current research**

Thus, although the adult literature shows that a shared experience changes adults’ attitudes to the content of that experience, and the developmental literature suggests that using shared cognition while engaging in an activity might affect
children’s processing of that activity, we know little about whether children have a preference for engaging in a shared experience through joint attention, or whether this joint attention shapes children’s attitudes towards the content of that activity. In the current study, we aimed to separately measure children’s preference for engaging in an activity in joint attention and their attitude towards the content of that activity itself, outside of its initial social context. To do so, we conducted a study comprising of two separate sessions. In the first session (content attitude), children were shown a moving toy while sitting next to an experimenter who was either watching that toy as well or reading their own book. Next, the experimenter left the room and the child was given the toy, after which we measured how long the child continued to play with that toy while distractor toys were presented at set intervals (Chapter 6, Study 1). In the second session (social preference) children were watching a video while sitting next to an experimenter who was either watching the video with them, or read their own book. In this case, we presented distractor toys during the video (i.e., during the shared experience itself) and measured when the child left their seat to go play with the distractor toys (Chapter 6, Study 2). In both sessions, we then compared the joint and disjoint attention condition in terms of (1) how long it took for the children to leave the moving toy/video for the first time to explore the distractor toys and (2) how long they were engaged with the moving toy/video in total.
The development of the Liking Gap

The ontogeny of concerns about making a good impression after a shared experience.

As stated earlier, engaging in social interactions and forming and maintaining social relationships is crucial for human physical and mental wellbeing (Baumeister & Leary, 1995). Conversely, being ostracized and excluded from social interactions is a major source of stress and anxiety (K. D. Williams, 2007). As such, social interactions and shared experiences with novel individuals are exciting opportunities to satisfy one’s need to belong, but also come with a concern about one’s reputation and the impression one makes on others. Recently, researchers uncovered a peculiar social illusion that seems to be a direct manifestation of these concerns: after a short shared experience (i.e., a conversation) between two strangers, both individuals think their partner evaluates them more negatively than they evaluate their partner, a phenomenon referred to as the Liking Gap (Boothby et al., 2018).

Little is known about when this Liking Gap emerges in childhood, or how it develops. A plausible hypothesis, however, is that its development coincides with the development of children’s concern for their reputation and the impression they make on others. Research has shown that such concerns emerge around age 5 (Buhrmester et al., 1992; Engelmann et al., 2012; Engelmann & Rapp, 2018; Fu, Heyman, Qian, et al., 2016; Fujii et al., 2015; Leimgruber et al., 2012; Silver & Shaw, 2018) and that children’s understanding of reputations and impression management strategies becomes more
sophisticated throughout middle and late childhood (Aloise-Young, 1993; Banerjee, 2002; Banerjee et al., 2012; Banerjee & Yuill, 1999; Fu, Heyman, Qian, et al., 2016; Herrmann et al., 2019; Shaw et al., 2014; Watling & Banerjee, 2007).

**Current research**

To find out whether the Liking Gap follows a developmental trajectory similar to that of general concerns for reputation and impression management, we conducted a Liking Gap study with pairs of children between 4 and 11. In this study, children first engaged in a shared experience with a peer for five minutes, during which they built a tower together. Next, the children were separated and asked several questions about (1) how they felt about their partner and (2) how they thought their partner felt about them. We then examined at which age children’s answers to these sets of questions started to diverge.
Chapter 2. Watching a video together creates social closeness between children and adults¹

Introduction

One important and unique way through which humans create and maintain social relationships is by actively seeking out social activities in which they share mental states with others (e.g. attention, goals, or emotions) towards external stimuli. To feel closer to other individuals we gossip about third parties (Bosson et al., 2006; Dunbar, 2004) and share thoughts, preferences, and ideas about many other aspects of our lives (Aron et al., 1997). Although such forms of sharing are simple and straightforward for adults, young children may lack the language proficiency to engage in this kind of sharing. However, humans also seem capable of connecting with others by sharing experiences without the use of language, for example by coordinating vocalizations and movements in song and dance (Pearce et al., 2015; Tarr et al., 2015, 2016) or playing team sports together (Artinger et al., 2006). Thus, if adults can connect through shared experiences in which no language is required, then they might provide young children with an alternative vehicle to connect with others.

The best known (or at least most commonly studied) phenomenon in which infants share experiences with others is joint attention (Bakeman & Adamson, 1984; Carpenter, Nagell, & Tomasello, 1998). Joint attention is crucial for children’s

¹ This chapter is published in the Journal of Experimental Child Psychology (Wolf & Tomasello, 2020a)
development in a variety of social-cognitive domains. It plays a key role in social learning and cultural transmission (Carpenter & Tomasello, 1995; Cleveland et al., 2007), as well as in developing the capacity to coordinate behavior with others collaboratively through establishing shared goals (Tomasello & Hamann, 2012). However, the socio-emotional consequences of this process, specifically its role in the formation of social relationships, have not yet been studied in young children.

In research with adults, Wolf et al. (2015) found that even after simply watched something on a screen together, without otherwise interacting, participants reported feeling more positive about and socially closer to their partner. Similarly, joint attention might allow children to create closeness between them and the adults around them. Although this has never been studied experimentally, previous correlational studies have found that the tendency to initiate joint attention in typically developing children as well as children with Autism Spectrum Disorder (ASD) is correlated with parent reports of a sense of relatedness to their child (Mundy et al., 1994). Furthermore, some research examining the affect displayed during joint attention episodes has suggested that joint attention is related to positive intersubjectivity and a sense of relatedness between a child and their interaction partner (Mundy et al., 1992).

If joint attention is indeed a mechanism for infants to increase social closeness with others, this would help to explain why young children spend so much time trying to actively share attention with adults (e.g. by pointing and vocalizing) in contexts
where it is unlikely that such behavior would yield any instrumental benefits (Liszkowski et al., 2004). In other words, it is possible that children’s attempts to establish shared attention with an adult to something interesting seemingly just for the fun of it are in fact motivated by the desire to create or reinforce positive social relationships with their communicative partners. The aim of the current experiments is therefore to examine whether sharing experiences through joint attention is a mechanism through which children create social closeness with adults.

Over the years, different operationalizations of joint attention have emerged, each of which would require a slightly different experimental set-up. In developmental psychology, the classic operationalization of joint attention is derived from observations of triadic interactions between children and adults in which they attend to the same thing and acknowledge that they are doing so by, for example, looking to the partner during the interaction (Bakeman & Adamson, 1984; Charman et al., 1997; Charman, 2003; Itakura, 2004; Landry, 1995; Leavens & Bard, 2011; Leavens & Racine, 2005; Mundy, 2006; K. D. Williams, 2007). In line with this conceptualization, Study 1 examined the effect of joint attention on subsequent social closeness in a natural, minimally interactive setting. Specifically, we examined if children would be more comfortable approaching an adult after having watched a video with them, relative children in a control condition, who watched the same video while the experimenter sat next to them privately reading a book. In this study, as in real life, the children were free
to engage the experimenter about the video, who in those cases responded with an acknowledging, friendly look.

More recently others have taken a more cognitive perspective, in which joint or shared attention is based on inferences made about whether or not we are attending to something together (Baron-Cohen, 1995; Bayliss et al., 2013; Gomez, 2005; Grossmann & Johnson, 2010; Nuku & Bekkering, 2008; Peacocke, 2005; Shteynberg, 2015, 2018; Striano & Stahl, 2005) and/or to what extent we are aware of others making such inferences (Behne et al., 2012; Carpenter et al., 1998; Grossmann et al., 2013; Liszkowski et al., 2004, 2008; Siposova & Carpenter, 2019). In line with this line of reasoning, and to make sure that the effects in Study 1 were not simply due to the way the experimenter responded to the child after a joint attention bid, we conducted an additional study in which we examined whether a similar effect can be found in a non-interactive context. Thus, in Study 2, the adult was instructed not to respond to joint attention bids from the child, while being seated a little more to the front, becoming less available for joint attention bids, but still visible enough for the child to infer that they were attending to the same thing.
Study 1

Methods

Participants and design

In total, seventy-four 1.5 year olds and seventy-two 2.5 year olds from mostly middle-class Caucasian families living in a middle-sized US city participated. Participants came into the university lab to participate in a study employing a between subjects (Watching together vs. Control) design. Of the original sample, 9 participants were excluded from analysis because of errors in the procedure (e.g. experimenter errors, parent errors). Another 16 participants were too fussy during the manipulation and, therefore, were excluded from analysis (e.g. they did not engage in the manipulation). Finally, 25 participants did not complete the final measure (e.g. they were fussy, or had no interest in taking part in the dependent measure). The final sample for analysis therefore consisted of forty-eight 1.5 year olds (age range = 17.13 months to 18.87 months, $M_{age} = 18.15$ months, $SD_{age} = .55$ months) and forty-eight 2.5 year olds (age range = 30.60 months to 35.90 months, $M_{age} = 33.78$ months, $SD_{age} = 1.41$ months). This means that each condition had 24 participants per age group, 12 of which were male and 12 of which were female. All procedures in Study 1 were approved by the institute’s IRB.
Procedure

When parents arrived in the waiting room, Experimenter 1 (E1) gave the parent some background and instructions regarding the study and went over the informed consent forms. Next, E1 warmed up with the child using a marble run. E1 stopped warming up if the child voluntarily came over to E1 and took the ball to put in the marble run. All children in the final sample ended up taking the ball from E1. After the warm up E1 took the parent and the participant to the experimental room. Here, the child sat down in front of the screen. For the 1.5 year olds, we asked the parents to choose if the child would be more comfortable in a toddler chair, a baby chair with a table, or a pillow. The 30 month olds all sat in a toddler chair. In both age groups, the parent was seated on a pillow behind the child filling out forms.

When the child and parent had sat down, E1 left the room to get one of nine different female experimenters (E2). All experimenters participated roughly the same amount of times in each condition (with a maximum of one more in one condition than the other). E1 and E2 then came in together. E2 sat down on a chair next to the participant without introducing herself and put a book open on her lap. E1 then said that he/she would start an animal video on the screen. Next, E1 walked behind the curtain and started a silent video consisting of eight 15 second video clips of animal behavior.
In the watching together condition, both the child and E2 could see the screen and looked at the video together. In this condition, E2 ignored the book in her lap. In the Control condition, E2 could not see the screen because her visual field was blocked by a piece of cardboard (50 x 50 cm) attached to the screen. Thus, in this condition E2 did not watch the video together with the participant and instead looked at the book in her lap. As we initially wanted the procedure to feel like a natural interaction E2 was instructed to, in both conditions, acknowledge every joint attention bid (e.g. a look or a point) by looking back at the child with a neutral/friendly expression. This dynamic is comparable to the checking back during joint attention episodes (e.g. Scaife & Bruner, 1975). During the manipulation, whenever a child looked back or attempted to otherwise engage the parent, the parents were instructed to acknowledge the child, say that they were busy, and go back to the form. The parent occasionally intervened, always on instruction from E1, for example by putting the participant back into the chair if he or she started walking around.

When the video was over, E1 came out behind the curtain. At the same time, E2 got up and walked to the other side of the room. E1 asked the parent to turn around to face E2 and asked the child to move in front of the parent, behind a start line, facing E2. The moment the participant had moved to the line and was facing E2, E1 gave E2 a cue and started a stopwatch. E2 responded by taking out a stuffed animal from under a blanket and offering it to the participant. The moment the toy was offered E1 told the
participant “If you want, you can go play with the toy...”, which the parent repeated after E1 following the instructions received before the warmup phase. If the child had not approached after 15 seconds E1 gave E2 another cue to give the stuffed animal a hug, and offer it again. After another 15 seconds, E1 gave a cue to E2, who responded by putting away the stuffed animal. She then took a toy truck with blocks from underneath a blanket and offered one of the blocks to the participant. Again, E1 cued E2 after 15 seconds. In this case, E2 tilted the back of the truck so that the blocks would fall out. She then grabbed another block and offered it to the participant. Finally, if the participant had still not approached, E1 cued E2 to put the truck away and take a marble run identical to the one in the warmup phase. E2 put the marble run next to her and offered a ball to the participant. During this part of the experiment, E2 looked at the participant, but never said anything to them.

**Measures and coding.**

Two cameras recorded the entire experimental session (i.e. manipulation and final measure). After each session, the research assistant that was part of the procedure split up into a manipulation video file and a dependent measure video file. A research assistant that was not part of the session then first checked the manipulation and then the dependent measure for exclusion criteria (e.g. parent or experimenter error, child fussiness). This procedure made sure that exclusion for behavior during the manipulation was uninformed by the dependent measure. Next, a different research
assistant coded the manipulation for looks to E2 and bids (i.e. a point or vocal expression accompanied by a look at E2), and the dependent measure video for approach latency. Approach latency was defined by the time the child took from the moment the participant was in place and stuffed animal was held out for the participant to take, to the moment the participant first touched one of the toys. A second coder coded 25% of all trials for latency, looks and bids for coder reliability (all Intraclass correlation coefficients: $r's > .90$).

**Results**

**latency**

As the residuals of the latency data for each group were not normally distributed (Kolmogorov-Smirnov test and Shapiro Wilk test both $p's < .001$) and several types of data transformation did not resolve this issue, non-parametric statistics were warranted.

As during the experiments, the behavior of the 18 month olds suggested that the procedure was too difficult for them. They were often fussy, did not want to sit in the chair and watch the video, and seemed to have trouble keeping track of who was watching what. This suspicion was confirmed by the approach latency results: An overall Mann-Whitney test across both age groups showed no difference between conditions (Mann-Whitney: $p = .097$; see Figure 1). The 1.5 year olds in the Joint Attention condition approached more or less equally fast ($M = 23.14, SD = 20.10$) as the 1.5 year olds in the Disjoint Attention condition ($M = 24.36, SD = 19.61$).
For the 2.5 year olds, we did find an effect of the manipulation on approach latency (Mann-Whitney: $p = .048$, $r = .286$). The 2.5 year olds approached E2 faster after the Joint Attention manipulation ($M = 17.07$, $SD = 16.91$) than after the Disjoint Attention manipulation ($M = 29.46$, $SD = 22.19$; see Figure 2).
Figure 2: Distribution of approach latency for 2.5 year olds in the watching together versus control condition (study1)

Looks and Bids.

For both looks and bids, the residuals of the data were not normally distributed (Kolmogorov-Smirnov test and Shapiro Wilk test for all variables in both age groups $p$’s < .005). Mann-Whitney tests showed no differences between conditions in either age group (all $p$’s > .05). The means and standard deviations of all variables in Study 1 can be found in Table 1 and 2.
Table 1: Means and standard deviations of latency, looks and bids per condition for the 1.5 year olds in Study 1.

<table>
<thead>
<tr>
<th></th>
<th>Joint Attention</th>
<th></th>
<th>Disjoint Attention</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Latency</td>
<td>23.14</td>
<td>20.10</td>
<td>24.36</td>
<td>19.61</td>
</tr>
<tr>
<td>Looks</td>
<td>7.21</td>
<td>3.13</td>
<td>6.83</td>
<td>3.24</td>
</tr>
<tr>
<td>Bids</td>
<td>2.33</td>
<td>2.99</td>
<td>2.13</td>
<td>3.83</td>
</tr>
</tbody>
</table>

Table 2: Means and standard deviations of latency, looks and bids per condition for the 2.5 year olds in Study 1.

<table>
<thead>
<tr>
<th></th>
<th>Joint Attention</th>
<th></th>
<th>Disjoint Attention</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Latency*</td>
<td>17.07</td>
<td>16.10</td>
<td>29.46</td>
<td>22.19</td>
</tr>
<tr>
<td>Looks</td>
<td>3.67</td>
<td>3.06</td>
<td>4.21</td>
<td>2.96</td>
</tr>
<tr>
<td>Bids</td>
<td>1.29</td>
<td>2.66</td>
<td>.21</td>
<td>.59</td>
</tr>
</tbody>
</table>

(* indicates significance at the .05 level)

**Frequency of children not completing the measure.**

There were many reasons for different children not completing the final measure (e.g. fussiness, a lack of interest in the dependent measure, sometimes caused by the participant wanting to watch more videos). When we examined the patterns of non-completions across conditions we found no condition difference in non-completions in the group of 1.5 year olds (7 participants vs. 8 participants, $z = -.240, p = .81$). However, we did find an unexpected difference in non-completions for the 2.5 year olds (10 participants vs. 0 participants $z = 3.023, p = .003$).
Discussion

The results of Study 1 show that 2.5 year olds, but not 1.5 year olds approach an experimenter faster after having been in joint attention with that experimenter than when the child and E2 have been attending to different things. However, the effect in Study 1 is only barely significant. Furthermore, the fact that there was an unexpected condition difference in the amount of children that did not complete the dependent measure in the group of 2.5 year olds suggests that a closer look to the experimental paradigm was necessary.

During data collection, the way in which some children reacted to E2 responding to their looks and bids by looking back raised concerns about the role of that allowing the experimenters to respond to the looks and bids of the child might have played in relation to their subsequent approach latency. First, we sometimes noticed children responding fearful to E2 looking back. This suggests that some of the looks towards E2 might have been explorative (i.e. trying to figure out who this stranger is that did not say a word and just sat next to them) rather than communicative (i.e. trying to share attention). As some children become more fearful after being addressed directly by a stranger, it could be that for some children in the sample the response to their looks might have worked counterproductively, eliciting anxiety rather than a sense of shared experience.
Second, we noticed that even in the Control condition some children still pointed or made bids for attention while looking at E2. As E2 responded by looking back (i.e. similar to what one might expect in a joint attention interaction), some children in the Control Attention condition might thus have used this social engagement to wrongfully infer that E2 could see/had been looking at the video with them. Furthermore, even if children did not erroneously infer joint attention based on E2 responding to their looks or bids, this social engagement would still have had an effect, adding noise to the data.

To address these issues, we conducted an additional study in which we set out to replicate the effect found in Study 1 using a procedure in which experimenters did not respond to looks or bids of the participant. One additional advantage of this new procedure is that it is a more direct replication of the Wolf et al. (2015) procedure used with adults, in which participants were instructed not to communicate with each other, as well as the more cognitive conceptualization of joint attention (Baron-Cohen, 1995; Bayliss et al., 2013; Gomez, 2005; Grossmann & Johnson, 2010; Nuku & Bekkering, 2008; Peacocke, 2005; Shteynberg, 2015, 2018; Striano & Stahl, 2005). Finally, as mentioned earlier, the procedure seemed too demanding for the 1.5 year olds and we therefore decided to only run this new, cleaner experimental procedure with a group of 2.5 year olds.
Study 2

Methods

Participants and design

In Study 2, a sample of Seventy-four 2.5 year olds from the same population as Study 1 participated in a design identical to that of Study 1. Of the original sample, 7 participants were excluded because of procedural errors, 10 participants were excluded for fussiness and 9 participants were excluded because they did not complete the final manipulation. For 1 participant there was a discrepancy in birth date between the birth records and the details the parent provided. As such, this participant might have been a couple of days older than 36 months. The final sample therefore consisted of forty-eight 2.5 year olds (age range = 30.64 months to 36.36 months, $M_{age} = 33.79$ months, $SD_{age} = 1.66$ months) who were equally distributed over conditions. Like in Study 1, gender was equally distributed and counterbalanced over conditions.

Procedure and measures.

The procedure of Study 2 was identical to that of Study 1 with two subtle changes in the manipulation. In Study 2, E2 was now instructed not to respond to the child at all. Also, in comparison to Study 1 the chair of E2 was moved towards the screen 10 cm while the participant’s chair was moved back 10 cm. This made E2 slightly less accessible for the participant to engage E2, as the child was now outside E2’s visual field.

---

2 This study was an honors thesis project by Rachel Galagos
In Study 2, the coder reliability of the latency, looks and bids was similar to that of Study 1 (all Intraclass correlation coefficients: $r$’s > .95).

**Results**

**Latency**

Similar to in Study 1, the residuals of the time it took children to approach were not normally distributed and data transformations did not resolve this issue, warranting non-parametric statistics. A Mann Whitney U test showed that there was an effect of condition on approach latency (Mann-Whitney: $p = .029$. $r = .315$). Participants in the Watching Together condition approached faster ($M = 18.02$, $SD = 20.09$) than participants in the Control manipulation ($M = 26.86$, $SD = 21.62$; See Figure 3).
Figure 3: Distribution of approach latency for 2.5 year olds in the watching together versus control condition

Looks and Bids

Due to technical problems, the video material needed for coding the looks and bids was lost for one of the participants; this data was therefore coded as missing. Like in Study 1, the residuals of the looks and bids variables were not normally distributed and we therefore used non-parametric statistics. Mann Whitney U tests did not yield a difference between conditions for looks and bids (all $p$’s > .05). The means and standard deviations of all variables are displayed in Table 3.
Table 3: Means and standard deviations of latency, looks and bids per condition for the 2.5 year olds in Study 2.

<table>
<thead>
<tr>
<th></th>
<th>Joint Attention</th>
<th>Disjoint Attention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Latency*</td>
<td>18.02</td>
<td>20.09</td>
</tr>
<tr>
<td>Looks</td>
<td>4.42</td>
<td>2.17</td>
</tr>
<tr>
<td>Bids</td>
<td>.54</td>
<td>.98</td>
</tr>
</tbody>
</table>

(* indicates significance at the .05 level)

**Frequencies of non-approaches**

Importantly, a z-test for proportions showed that there was no difference ($z = -.319$, $p = .75$) between the proportion of children that did not approach in the Watching Together condition (4 out of 52) and the Control condition (5 out of 53).

**Discussion**

Study 2 replicated the effect of engaging in a shared experience on approach latency found in Study 1. Although the mean difference is smaller in Study 2 than in Study 1, the effect size is slightly larger (i.e. $r = .286$ for Study 1 and $r = .315$ for Study 2), suggesting that there might be slightly less noise in the data (although effect size differences as small as this one should be interpreted with caution). Most importantly, Study 2 was successful in getting rid of the unexpectedly skewed distribution of children that did not complete the final measure in Study 1. As such, Study 2 corroborates the findings in Study 1, suggesting that 2.5-year-old children socially connect with adults through shared experiences.
General discussion

Together, these two studies provide support for the notion that sharing experiences through joint attention in both a minimally interactive as well as a non-interactive context influences how 2.5 year olds respond to novel adults. More specifically, the studies show that 2.5 year olds feel more comfortable with an adult after having watched a video together with that adult than when an adult was sitting next to them reading a book. These results imply that sharing experiences with others may play an important role in their developing social relationships.

Study 1 showed neither an effect nor a trend for the 1.5 year olds. One consideration with regard to the sample of 1.5 year olds concerns their visual perspective taking capabilities. Although 1 and 1.5 year olds have been shown to process that others can see things they cannot see (Moll & Tomasello, 2004), understanding that others cannot see what they can see might be more difficult. As such, the 1.5 year olds might have had trouble understanding that E2 in the control condition could not see the video, which might partially explain why they made joint attention bids even when E2 could not see what the 1.5 year olds wanted them to see. However, during the experiment E2’s gaze was not directed towards the screen, but instead slightly downwards towards their own book. Thus, although the 1.5 year olds might have erroneously inferred that E2 had visual access to the video in the control condition, it seems less likely that the children inferred that E2 was actually watching the video at the
same time they were. As such, it seems unlikely that the current lack of results for the 1.5
year olds can be completely attributed to their lack of visual perspective taking skills.
However, as stated previously, it was also clear that the 1.5 year olds in this study had
trouble sitting still to watch a video without sound for 2 minutes, especially without
trying to engage their parents in the interaction. The videos without sound might simply
not have been arousing enough to keep their attention (relative to the novelty of the
room and the experimenters). We can therefore not conclude that the lack of results for
the 1.5 year olds reflects a fundamental difference in infants’ social development. More
research using less demanding procedures is necessary to provide a more definitive
answer to that question.

The results for the 2.5 year olds are in line with previous effects found with
adults (Wolf, Launay, et al., 2015) and, more recently, also in great apes (Wolf &
Tomasello, 2019). However, recent studies with adults have shown that the extent to
which this effect occurs might be moderated by the content of the stimulus. First, the
social bonding properties of social bonding have been shown to be contingent on the
degree to which a stimulus is arousing and/or its valence (Rennung & Göritz, 2015).
Second, Haj-Mohamadi, Fles, & Shteynberg (2018) found that the degree to which the
content of a stimulus is in line with prior beliefs moderates the social bonding effect of
joint attention, especially for belief confirming or disconfirming stimuli that tap into
beliefs that are firmly rooted into one’s social identity or group membership, such as
beliefs in evolution or creationism (Haj-Mohamadi et al., 2018). Being exposed to stimuli that go against such deeply entrenched beliefs might decrease affiliative social processes, which is mentioned by Haj-Mohamadi et al. (2018) as an explanation for why they only found a social bonding effect of joint attention between participants when all the stimuli they had been exposed to during the manipulation affirmed their beliefs.

This raises the question how what exactly the underlying psychological mechanics of the social bonding effect of joint attention are. Previous research has shown that shared experiences are more elaborately processed and deeply encoded in memory, intensifying the emotional experience of a stimulus, thereby shaping one’s attitude towards it (Boothby et al., 2014; Eskenazi et al., 2013; Shteynberg, 2010, 2010; Shteynberg, Hirsh, Apfelbaum, et al., 2014). As such, a mildly pleasant experience, such as watching a video, might become more pleasant by doing it together. This increased pleasantness, in turn, might become associated with the individual with whom the video was watched together, causing a greater willingness to interact afterwards. On the other hand, when using a more cognitively rich conceptualization of joint attention (e.g. Carpenter & Tomasello, 1995; Tomasello, 1995) one might emphasize the creation of common knowledge or common ground between individuals through a joint attention interaction, making subsequent interactions both smoother and more meaningful (Clark et al., 1983), and thus more appealing to seek out again in the future. These two mechanisms are not mutually exclusive, and more research is needed to find out if and
how these mechanisms individually (and/or simultaneously) facilitate the creation of social closeness between individuals.

It is important to note that the type of social connection described in the current study is different from classic attachment dynamics (Bowlby, 1982). Attachment prototypically concerns primary caregivers and focuses on issues of danger-safety and availability-abandonment. It is particularly salient in response to unfamiliar environments, which might contain potentially dangerous agents or objects. The results of Study 1 and 2 imply that children also socially connect with even novel adults in a way that is not driven by attachment concerns, narrowly defined. In contrast, social relatedness or closeness through shared experiences seems to stem from a fundamental motivation to engage in triadic interactions that provide opportunities to share internal states and activities with others (Tomasello, Carpenter, Call, et al., 2005).

Such a motivation might explain why young children already actively engage in triadic interactions so often, even when their behavior does not seem motivated by any particular instrumental goal. Research on (proto) declarative and expressive pointing in twelve- and eighteen month olds has shown that children point specifically to share attention with adults even in the absence of an instrumental goal (Liszkowski et al., 2004). Yet, the children in the Liszkowski et al., (2004) study were not satisfied when an experimenter just looked at the child and expressed emotions to them; they were not simply trying to engage the adult in any form of interaction. Instead, the experimenter
had to look at where they pointed and then look back to acknowledge that they had seen it for the child to seem satisfied. So if the pointing gesture was not motivated by a desire to achieve an instrumental goal nor by a desire to simply be socially engaged, then what were those children trying to achieve? The current research might help to provide an answer to this question, suggesting that such declarative/expressive pointing gestures might be motivated by a more specific desire to engage in triadic interactions for the purposes of creating a social connection with others, for example when the willingness and/or opportunities to engage in physical interactions are limited.

This might be particularly relevant when looking at the development of adult-child dynamics in our evolutionary history. That is, the results of the current study might hint at an evolutionary mechanism that allows children to elicit solicitude in cooperative breeding groups, in which multiple children who have started weaning (like the 2.5 year olds in Study 1 and 2) are typically looked after by only a few adults, and therefore have to compete over the attention of these adults (Burkart et al., 2009a; Hrdy, 2007a). It has been suggested that in such an environment, children who are successful of connecting with adults from a distance are more likely to acquire the care and attention they need, providing an evolutionary advantage for those children who develop their shared cognition earlier relative to those in whom such capabilities emerge later. This, in turn, might help explain why shared cognition emerges so early in
ontogeny, in particular before children have developed any of the other necessary (e.g. physical) capacities for successful collaboration in the adult world (Tomasello, 2019).

One important question to consider is how the socio-emotional effect of a shared experience compares to the social closeness felt in a long lasting social relationship. The measures in the current studies were aimed at capturing a degree of comfort and positive attitude towards that individual, as well as a willingness and motivation to interact with them. There is no doubt that these aspects are core components of any social relationship. However, close social relationships are also characterized by, for example, an expectation to reciprocate social support when needed. Such expectations do usually not emerge after a single interaction, but take time and repeated interactions to develop. As such, creating social closeness through shared experiences might be relatively more important in the early stages of relationship formation, before factors such as social support start to play a role.

Generating a better understanding of the role that triadic interactions play in children’s social development might generate new insights into how social networks are formed during the earlier stages of human ontogeny. Although the scope of the current research was to examine the effect of shared experiences on social closeness between children and adults, there is no reason why this mechanism might not work for interactions between peers. It is worth noting that, compared to their interactions with adults, infants rarely engage in joint attention with other infants (Bakeman & Adamson,
1984; Chen et al., 2009), only beginning to do so regularly as their peer interactions ramp up in importance after three years of age. Nevertheless, studying shared experiences between young children might illuminate how children’s social world transitions from a parent/adult dominated social world to a social world in which peer relationships start to play an increasingly more important role.

Finally, the results of the current study also bear implications for research on and clinical practices with children with ASD. If young children connect with others through shared experiences then impairments in shared cognition, such as those found in children with ASD (Charman et al., 1997), will have profound consequences for their socio-emotional development. As such, the current research help inform research into the socio-emotional development and social relationships of children with ASD. This knowledge might then be used to inform clinical interventions aimed at teaching high functioning children with ASD new ways of connecting with others, perhaps through behaviorally grounded (rather than communicative) triadic interactions that they find more natural. This might help children with ASD to reduce the relatively higher level of loneliness and poorer quality of friendships they report compared to typically developing children (Bauminger et al., 2003; Bauminger & Kasari, 2000).

Overall, the current series of studies provides a new perspective on the role of shared experience in children’s social development. Understanding the socio-emotional value that shared experiences have for children might help explain a variety of their
social and communicative behaviors that have been the focus of cognitive study for decades. Most importantly, the present research sheds new light on the development of humans’ remarkable motivation to invest much time and energy in order to share our mental states with others to specific parts of the world around us.
Chapter 3. Visually attending to a video together facilitates great ape social closeness

Introduction

Humans create and maintain social relationships in ways that are seemingly unique in the animal kingdom. Specifically, humans are able to create social closeness through all kinds of shared activities and experiences that do not require direct physical interaction, but instead seem to satisfy a fundamental need to share experience with other individuals (Echterhoff et al., 2009). Although the precise psychological mechanisms through which such activities result in social closeness remain unclear, humans have been shown to connect with one another by doing such things as making music together (Pearce et al., 2015), acting together in synchrony (Hove & Risen, 2009), dancing together (Tarr et al., 2015, 2016), playing team sports together (Artinger et al., 2006), or by sharing experiences through gossip (Dunbar, 2004) or attitudes (Bosson et al., 2006), or disclosing personal information (Aron et al., 1997). In a recent study, Wolf et al. (Wolf, Launay, et al., 2015) demonstrated that even after a minimal shared interaction in which participants were attending to the same thing without otherwise communicating, they reported feeling closer to that participant (Rennung & Göritz, 2015).

Throughout the animal kingdom, the individuals of many species act in coordination with conspecifics. For example, dolphins often behave in synchrony (Fellner et al., 2006), many bird species coordinate their song and dance in a mating context.

---

3 This chapter is published in the Proceedings of the Royal Society B: Biological Sciences (Wolf & Tomasello, 2019)
(Ullrich et al., 2016; H. Williams, 2001), and great apes travel together (Symington, 1990) and sometimes hunt monkeys together (Boesch, 1994). But do behavioral interactions in which to individuals focus on an external stimulus together create stronger social relationships or bonds between participants? To our knowledge, there are no studies examining such a relationship in any nonhuman species, and indeed some theorists have suggested that this method of social bonding might be uniquely human (Tarr et al., 2016; Wolf, Launay, et al., 2015).

As always in comparison to humans, great apes are a special case because of their close phylogenetic connection. Operational definitions of social closeness (bonding) in great ape research usually rely on interactions involving physical closeness (e.g. grooming and physical play (Behncke, 2015; Hare et al., 2015; Lehmann et al., 2007) and/or spatial proximity (Kasper & Voelkl, 2009). However, given that apes do engage in a variety of coordinated, and even to some degree cooperative activities, such as building and fighting in coalitions and alliances (Harcourt & De Waal, 1992), as well as traveling and hunting in groups (Boinski & Garber, 2000), the question is whether, like humans, great apes have evolved a psychological mechanism that leads them to create social closeness with others through shared experiences. On the other hand, it might be that connecting with others through shared experiences is a uniquely human phenomenon.

To answer this question, we adapted the Wolf et al. (2015) paradigm for apes and conducted two studies in which participants shared the experience of attending to a video together with a human experimenter (Study 1) or a conspecific (Study 2). In the control condition, a human experimenter (Study 1) or conspecific (Study 2) sat in the same place
but was not watching the video. We then compared the apes’ subsequent behavior towards their partner – approaching and/or remaining in physical proximity - between the two conditions.

**Study 1**

Methods

Participants and design

Nineteen chimpanzees (*Pan Troglodytes*) and seven bonobos (*Pan Paniscus*) at the Wolfgang Kohler Primate Research Center (WKPRC) in the Leipzig Zoo participated in a two condition (Watching together (WT) vs. control) within subject experiment. All of them had previously engaged in social and cognitive experiments with humans. For two chimpanzees one of the trials yielded uninterpretable behavior\(^4\) and were therefore excluded from analysis. The final sample for which we collected behavioral data thus consisted of seventeen chimpanzees (Mean age = 26.4 years, 7 males) and seven bonobos (Mean age = 18.6, 3 males). Additionally, we also collected eye tracking data for fifteen chimpanzees (for two chimpanzees eye tracking data was not collected due to their physiology\(^5\) interfering with the eye tracking set up).

\(^4\) The videos showed that in these trials the participant did not enter the area in front of the experimenter, but merely walked by. However, the experimenter had interpreted that behavior as an approach and had therefore stopped the trial, meaning that no definitive approach latency could be coded for these trials.

\(^5\) One chimpanzee (Jeudi) had a condition in one of her eyes which caused the eye tracker not to be able to track that eye. The other chimpanzee (Riet) had a sizable swelling in the right part of her upper lip which caused her to tilt her head to the side while drinking from the juice tube. This caused the eye tracker not to be able to track her right eye.
**Ethics statement**

Study 1 was approved by the Wolfgang Köhler Primate Center Animal Research Committee, and was done in accordance with all of the governing laws and regulations concerning research with animals in Germany. In addition, the WKPRC has additional animal health and safety standards.

**Set-up**

Participants were sitting in a booth, looking through a Plexiglas screen to a pc monitor placed in front them. Underneath the monitor there was an eye tracking camera. Outside the room were two laptops, one of which was used for displaying stimuli on the monitor that the participants were looking at, while the other one was controlling the eye tracker. All trials were videotaped using three cameras. Participants sat down in front of a Plexiglas screen through which a mouthpiece of a juice tube stuck out (Figure 4).

**Procedure**

The experimental coordinator (EC) set-up the room for the appropriate condition and turned on the cameras. Next, the EC filled the juice tube with diluted grape juice, waited for the subject to come in\(^6\), made sure the subject was drinking from the juice

---

\(^6\) All participants were used to being moved around by keepers. Sometimes they were motivated by placing some food in the area where they had to wait for the study to begin. Furthermore, all subjects in Study 1 had extensive experience with juice tubes and therefore required no training in using them.
tube and left the room. The EC then started the eye tracker and signaled to experimenter 1 (E1) that they could go in.

Participants engaged in two trials of the procedure. There were two female research assistants that assisted in the procedure, each of which the participants had never seen or interacted with before. Each experimenter only engaged with each participant once and all participants thus interacted only one time with each experimenter. The experimenters were counterbalanced across conditions to prevent potentially confounding effects of a systematic preference of the participants for one of the two experimenters. At the same time, the order of the conditions was also counterbalanced to prevent potential learning effects from biasing the results.

When E1 got into the room, they sat down next to the screen. In the WT condition, the screen was turned towards the experimenter so that both the participant and the experimenter could see the screen (Figure 4A). In the control condition, the screen was turned away from the experimenter so that only the participant could see the screen (Figure 4B).

After E1 had sat down, EC started one of several one-minute videos. These videos were excerpts from a longer video of a playing juvenile chimpanzee. This video was chosen based on a recommendation of researchers conducting studies with

---

7 The chimpanzees and bonobos participating in the study have extensive research experience (i.e. testing occurs daily, throughout the year, for over 20 years), and are therefore used to interacting with humans they have not encountered before.
chimpanzees and bonobos using video stimuli. The key consideration was that the videos should be (1) interesting enough for participants to sit down and attend to the video and (2) not so arousing that they would elicit stress and/or behavior that would cause them to disengage from the video. We were therefore advised to show participants a video of great apes that were not in their own group, which would certainly capture their interest. However, showing them a video of adult males and/or females might contain cues of dominance or mating, which would make the video too arousing. We therefore decided that a video of a playing juvenile chimpanzee would serve the purposes of the current study best.

![Watching together](image1) ![Control](image2)

**Figure 4:** Visual experience of the ape in the Watching Together condition (A) and the control condition (B). Note the eye tracker on the table in front of the television monitor.
**Eye tracking apparatus and usage.**

To measure eye movement, we used a 60 Hz; X120, Tobii infrared eye tracker (Tobii technology AB, Stockholm, Sweden), commonly used at the WKPRC. This eye tracking apparatus records eye movement in a non-invasive manner, without head restraint. Previous research in this facility has shown that the acrylic safety panels do not interfere with the eye tracking measurements (e.g. Krupenye, Kano, Hirata, Call, & Tomasello, 2016). Participants were kept relatively stationary by supplying them with diluted grape juice through a custom-made juice tube device, created from a medical drip that was fitted with a mouthpiece at the end of the tube, which fitted through a hole in the Plexiglas screen. The provision of juice was not contingent on the participants’ eye gaze.

The eye tracker was calibrated on the real world visual space of the participant. To do so, a separate camera was fixed on the glass above the participant. Furthermore, a black screen was placed behind the scene on which the eye tracker needed to be calibrated. This screen contained two marks which indicated the corners of the visual field of the camera. Each participant was calibrated by presenting them with physical objects (e.g. toy fruit) in those two corners. All calibrations were checked manually and replaced if necessary. For analysis, we created 2 areas of interest. One area of interest encompassed the area of the scene in which the screen was placed during the
manipulation. The other area encompassed the visual field of the participant in which the experimenter was sitting during the manipulation.

**Approach latency**

To get a behavioral measure of the attitude of the participants towards the experimenter we used approach latency towards that experimenter. Approach latency was operationalized as the participant crossing the area four tiles behind the Plexiglas screen behind which experimenter 2 was now sitting. A ceiling camera was used to determine if the participant stepped over that line. If the participant did not approach after 30 seconds, the experimenter started rattling the mesh to prompt the participant to come over. If they still had not approached after 60 seconds, the experimenter, in addition to rattling the mesh, also called out their name. All participants approached within 90 seconds. The moment the participant entered the approach area, the trial was finished.

**Results**

**Eye tracking manipulation check.**

The eye tracking data showed that all participants in all trials had attended the video during the manipulation (Mean looking time = 26.03, SD = 12.13) and that all of them had looked at the experimenter at some point after the experimenter had sat down (Mean looking time = 5.68, SD = 7.03). As the data was not normally distributed, we conducted a within subjects Wilcoxon signed rank test (two sided) on the 30 trials (two
times 15 subjects). We found no differences in the time the participants spent watching (1) the experimenter or (2) the screen between the WT condition and the control condition ($p$’s > .21).

**Approach latency**

Approach latency was coded using the ‘interact’ software package. A second coder double coded 50% of the trials (24 out of 48). The angle of the videos that were coded was such that it was impossible to see what condition that subject was in. Coder reliability was high (Intraclass correlation coefficient= .999). As the data were not normally distributed and data transformations did not resolve this issue, non-parametric statistics were warranted. To test if the manipulation had an effect, as well as whether this effect was different for the different species, we conducted a two (condition: watching together versus watching alone) by two (species: chimpanzee versus bonobo) non-parametric mixed model with approach latency as the dependent variable (Noguchi et al., 2012). Results showed no main effect of species on approach latency (modified\(^8\) ATS ($df = 1$, 10.90) = 3.98, $p = 0.07$), nor an interaction effect between condition and species (ATS ($df = 1$) = 0.01, $p = 0.92$). Crucially, we did find an effect of condition on approach latency (ATS ($df = 1$) = 5.61, $p = 0.017$; Figure 5). Participants from both species approached the experimenter with whom they had watched the video together faster ($M$

\(^8\) The between subjects factor in this model is evaluated with a modified ANOVA Type Statistic (ATS); see (Noguchi et al., 2012).
= 14.86, $SD = 13.18$) than the experimenter that had been reading her own clipboard ($M = 26.78, SD = 2.93$).  

Figure 5: Approach latency per condition (Study 1)

---

9 See Appendix A, Figure 20 for an overview of the individual response times per condition, grouped by species.
Discussion

These results suggest that visually attending to a stimulus together creates some kind of social connection between a great ape and a human, such that the ape experiences an increased motivation to approach the human experimenter. What is still unknown, however, is whether these results indicate a general psychological mechanism in great apes, or rather a particular way that apes in zoos relate to humans based on their extensive experience with them. Previous research has found that great apes are, for example, more likely to understand and produce an imperative pointing gesture while interacting with humans than when interacting with conspecifics (Hare et al., 2000; Leavens et al., 1999b; Lyn et al., 2010; Tempelmann et al., 2013). Similarly, the elaborate interactions with humans of the sample in Study 1, including frequent testing, as well as their dependence on humans for food might make them specifically attuned to what humans are looking at, and this might not generalize to interactions with conspecifics. In Study 2, therefore, we aimed at replicating these findings but between two chimpanzees (and with a different group of chimpanzees, that is, from the Ngamba Island Chimpanzee Sanctuary).
Study 2
Methods

Participants and design

Twenty-one chimpanzees (Mean age = 19.57, SD = 3.88, 10 males) at the Ngamba Island Chimpanzee Sanctuary (NICS) in Uganda participated in a study which had a design similar to Study 1. All of them were part of a single group and had previously engaged in social and cognitive experiments with humans before. In contrast to Study 1, the procedure of Study 2 required the participation of pairs of chimpanzees. The creation of the pairs was based on the judgement of the NICS keepers. We asked them to help us create a list with same-sex pairs of individuals who were likely to be tolerant of one another while being alone in the same room. In addition, we asked them to exclude pairs in which one of the individuals was likely to be scared or intimidated by the other individual and would try to hide or avoid their partner while being in the same room. Finally, we also decided that individuals should not go through the procedure more than four times (i.e. eight trials). Based on these constraints, we ended up with 36 pairs which we tested over 72 trials\(^{10}\) (one WT trial and one control trial per pair). (See Appendix A, Table 9 for an overview of how many times each individual was tested).

\(^{10}\) Due to technical problems, the videos of two trials were lost, and these trials were therefore replaced with a new trial. Excluding these pairs from the analysis does not meaningfully change the results of Study 2.
Ethics Statement

Study 2 was approved by the Ugandan Wildlife Authority (UWA) in accordance with governing laws and regulations concerning research with animals in Uganda.

Set-up

Participants were led into two adjacent rooms with a closed door between them. One room was smaller than the other, allowing us to seat participants in a 90-degree angle from each other (Figure 6, Appendix A, Figure 21). The participants sat down behind two Plexiglas screens, each of which had a juice tube mouthpiece sticking out from which they could drink diluted mango juice. In the WT condition, one computer monitor was placed in front the participants so that both participants could see the same screen (Figure 6A). In the control condition, two monitors were placed with a plastic barrier in between the screens, so that participants could not see each other’s screen (Figure 6B). In both conditions, pc monitors were placed so that the participants could still see their partner in his or her entirety. The pc monitors were connected to a laptop controlling stimulus presentation. Furthermore, the room in which the monitors were placed contained three cameras, which recorded the chimpanzees’ behavior during and after the manipulation.
Procedure: Individual exposure trial

Before we started any of the experimental trials, we wanted to expose the participants to the juice tubes. Furthermore, we wanted them to gain experience with the two-monitor set-up, because we wanted to make sure participants were aware that from the perspective of the different Plexiglas screens, it was impossible to see what was happening on the other screen (an inference which previous research has shown great apes are able to make; Hare, Call, Agnetta, & Tomasello, 2000). To do so, participants entered the rooms\textsuperscript{11} by themselves while the door between the rooms was open. To

\textsuperscript{11} As in Study 1, subjects in this group were used to being moved around by keepers, who sometimes used food to direct them towards the right location where they waited to be let into the testing enclosure. This is very similar to the strategy the keepers use every evening when trying to get chimpanzees in specific sleeping enclosures (to minimize conflict at night) by throwing small food pallets in specific areas. As such, the movement of subjects in Study 2 was in line with the chimpanzees' daily routine.
reduce the likelihood of carry-over effects from the exposure trials into the experimental trials, the stimulus for the exposure trials was an unrelated video of (non-chimpanzee) animal behavior used in a previous study with human children. All chimpanzees almost instantly understood the mechanics of the juice tube, and drank from them during the exposure trials.

The first phase of the exposure trial consisted of participants freely roaming around the two rooms for three minutes. In the second phase, we wanted to expose the participants to the juice tube and make sure they had experienced the perspective from both sides of the set-up (as some individuals merely sat down at the first Plexiglas screen they came across and did not roam around any further). In the second phase, we therefore offered them juice through the juice tube at the Plexiglas screen of the room they were not in at that moment. This way, all the apes walked around to the other room and thus saw the set-up from both perspectives at least once.

**Procedure: Experimental trial.**

After we had completed all the exposure trials, we started engaging pairs of participants in the experimental trials. All pairs engaged in a WT trial and a control trial, at least a day apart, with the order of the conditions counterbalanced across pairs. In both conditions, two keepers of the NICS let the two individuals into their individual rooms to engage in the manipulation. During the manipulation, the door between the rooms was closed. The moment the participants came in the experimenter turned on the
juice tube. In all cases, this motivated the participants to come to the Plexiglas screen and drink from the juice tube. The experimenter then started the same videos used in Study 1 (i.e. the videos of the playing juvenile chimpanzee) and left the room. When the video was finished, the experimenter came back into the room and took out the juice tubes. He then left the room, while at the same time one of the keepers opened the door between the rooms the participants were sitting in. Next, the experimenter and both of the keepers moved to a location where the participants would not be able to see them anymore. The participants were then left to do what they wanted for three minutes, during which their behavior was recorded.

Measures

As the manipulation in Study 2 was no longer an interaction between an ape participant and an experimenter who stayed in one place, but instead a spontaneous interaction between two apes, we could no longer use approach latency as a proxy for social closeness. Instead, we measured participants’ subsequent physical proximity as our main dependent variable, consistent with previous research on social networks in great apes (Kasper & Voelkl, 2009). As a secondary dependent measure, we also looked at interactive behaviors after the manipulation. In order to obtain sufficient behavior to

---

12 One of the keepers hid behind the enclosure, where he was able to look around the corner in case he heard aggressive vocalizations. This allowed him to intervene if, in his judgement, the fighting put one of the animals at risk for serious physical harm. In none of the three instances of fighting that occurred did the keeper feel the fighting was severe enough to intervene, and no injuries were sustained by any of the participants.
analyze in the absence of experimental prompts (as in the first study) we decided that a longer time window during which behavior was recorded was necessary. However, at the same time we also had to consider the possibility that the effect of the manipulation would wear off over time, potentially adding noise to the data. Additionally, the staff and researchers that had worked with these subjects before cautioned us that for some subjects, keeping them in their enclosure with nothing to do while part of their group members had already left for the forest would make them uncomfortable. Based on these methodological and animal welfare considerations, we decided to set the time window during which their behavior was measured to three minutes. After three minutes, data collection stopped, as the keepers and experimenter returned to the enclosure to let the subjects out.

**Proximity.**

Proximity was coded on three different levels. First, we coded the time the participants spent in the same room. However, as the surface of the big room in the enclosure was twice as large as the surface of the small room (see Appendix A, Figure 21), the base rate probability of a participant spending time in the big room was twice as high as the probability of them spending time in the small room. Furthermore, the rooms were several meters high, with hammocks hanging from the ceiling, where some of the apes spend at least part of the three minutes following the manipulation. As such, a proximity measure merely constrained by which room the apes were in did not
account for their actual physical proximity in three dimensional space. Also, it is important to note that the enclosures were surrounded by bars (and not opaque walls). This means that subjects could still see each other while being in different rooms, meaning that they could not simply hide away from their conspecific’s attention by going into a different room. As such, compared to actual physical proximity, being in a separate room, at least in this context, did not seem to be relevant for measuring social closeness.

To address these issues and reduce the amount of noise in the data, the main variable of interest was how much time participants spent in the same part of the room. That is, during coding, the big room was split up into two equal parts, causing the total surface area of the big and small room together to now be divided into three more or less equal spaces (see Appendix A, Figure 22). Furthermore, we only coded time spent in the same part of the room if participants were on the same level of the enclosure. This means that either both of them had to be on the floor, or both had to be hanging from the ceiling or sitting in a hammock attached to the ceiling in order for that time to qualify as time spent in the same part of the room.

Finally, we also coded the amount of time the participants spent at arm’s length. In these cases, participants were sitting so close to one another that they were either touching or would have been capable to touch each other if one of the individuals would
have stretched their arms. We also coded this for instances where the two participants were not technically in the same part of the room.

In addition to measures of physical proximity, we also measured the frequency with which two types of interactive behaviors occurred between participants. We coded the time participants spent grooming (i.e. sifting through the hair of another individual, see: Fellner et al., 2006) a common indicator of social affiliation in great apes (Hare et al., 2015; Lehmann et al., 2007). Additionally, we coded the time participants spent fighting (i.e. aggressively chasing or using physical aggression) with each other during the three minutes after the manipulation as a counter-indicative measure of social closeness. The amount of time individuals spend fighting within a dyad was subsequently subtracted from all the time individuals spend together in the proximity measures, as this was not the type of proximity that can be used as a proxy measure of social closeness.

**Results**

All behavior was coded with the BORIS software package for behavioral coding. In total, 25% of the trials (i.e. 18 out of 72) were double coded. Coder discrepancies were discussed and resolved, resulting in high coder agreement (Intraclass correlation coefficient $r = .995$). Next, we compared the participants’ WT trials to their control trials in terms of the time spent in each other’s proximity and the time engaged in interactive behaviors (see Table 4). As none of the data was normally distributed, we conducted a
within subjects Wilcoxon signed rank test for each variable (two sided) on the 72 trials (two times 36 pairs).

Figure 7: Time spent in the same part of the room (Study 2)
We found an effect of condition on the amount of time participants spent in the same part of the room ($p = .033$, $r = .251$, see Figure 7). Participants spent more time in the same part of the room in the WT condition ($M = 31.50$, $SD = 6.44$) than in the control condition ($M = 19.25$, $SD = 6.04$).

We found no effect of any of the other variables ($p's > .05$). As expected, the more general ‘in the same room’ measure seemed not sensitive enough to pick up a difference between conditions. Furthermore, proximity at arm’s length ($N = 18$), as well as fighting ($N = 3$) and grooming ($N = 1$) did not occur frequently enough to do statistical analyses on. With regard to grooming this is not surprising. As research has shown that chimpanzees only spend 6.8% of their daytime grooming (i.e. around 45 minutes a day; 20), a three minute window to measure grooming might simply have been too short for such behavior to occur frequently enough to make statistical inferences. However, as Table 4 shows, the trend of time spent (1) in the same room, (2) at arm’s length, (3) grooming, and (4) fighting showed a converging trend. Grooming only occurred after WT trials while fighting only occurred during Control trials. Furthermore, the condition means of the time participants spent in the same room and the time participants spent at arm’s length, despite lacking sensitivity or occurring too infrequent to analyze, show a trend similar to the effect found in the time spent in the same part of the room.
Table 4: Descriptive statistics of coded behaviors in Study 2.

<table>
<thead>
<tr>
<th></th>
<th>Watching Together</th>
<th></th>
<th>Control</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>N*</td>
<td>Mean</td>
</tr>
<tr>
<td>In same part of the room</td>
<td>31.5</td>
<td>38.61</td>
<td>25</td>
<td>19.25</td>
</tr>
<tr>
<td>In the same room</td>
<td>72.47</td>
<td>61.07</td>
<td>28</td>
<td>65.22</td>
</tr>
<tr>
<td>Arms-length</td>
<td>4.61</td>
<td>10.12</td>
<td>11</td>
<td>2.00</td>
</tr>
<tr>
<td>Grooming</td>
<td>1.13</td>
<td>6.75</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Fighting</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.03</td>
</tr>
</tbody>
</table>

* The number of trials in which this behavior occurred at least once.

**Discussion**

Extending the results of the first experiment, these results suggest that visually attending to a stimulus together with another individual elicits social closeness in great apes not only with human experimenters but also with conspecifics. Furthermore, the results were obtained from a different population of apes, living in a larger group, in an African sanctuary instead of a European zoo, suggesting that the effect of watching something together on great ape social closeness is not limited to apes living in a zoo environment.

**General Discussion**

The current results show that great apes behave more socially after an interaction in which they align their attention to an external stimulus. Study 1 showed that both chimpanzees and bonobos approach a human experimenter faster after having watched a video with them, suggesting that this effect can be found in the entire genus Pan. Study
2 replicated these findings in a different sample and extended them by showing that this effect is not limited to great ape’s interactions with humans, but also seems to occur in interactions between great apes. As such, the current findings shed new light on great ape social cognition and social behavior, as well as the evolutionary origin of connecting through shared experiences in humans.

Becoming socially closer to others through shared experiences such as dancing to music together or communicating about shared experiences has only been described in humans. It has therefore been suggested that this bonding mechanism is uniquely human, explaining (at least in part) why humans have larger social networks with more complex social relationships than other species (Dunbar, 2008, 2012). The current results imply, however, that some of the basic elements of this social bonding mechanism – eliciting social closeness by visually attending to something together with another individual – is present in humans through shared descent with other apes.

This is surprising because many researchers have argued that the capacity to experience reality as shared is uniquely human (Echterhoff et al., 2009; Tomasello, Carpenter, Call, et al., 2005), which implies that other species, including great apes, lack the psychological mechanisms to interpret an experience as shared and therefore do not engage in social activities solely for the purpose of generating social closeness. However, one can imagine that great-ape activities such as fighting together in a coalition or traveling together in a small group - based at least partly on visually attending to things
together – already elicit social closeness among ape individuals. This suggests that such activities aside from their instrumental purpose (e.g. traveling safely, acquiring or maintaining dominance in a group), might also function as a way to generate social closeness between the individuals partaking in the activity. As such, this psychological mechanism might be a previously unnoticed facilitator of great ape social relationships.

One must, however, be cautious when extrapolating a human shared experience-based bonding mechanism to great apes on the basis of the current studies. The current control condition for both studies was designed to keep all parts of the experience, aside from whether it was shared or not, as constant as possible, similar to studies on shared experiences in humans (Rennung & Göritz, 2015; Wolf, Launay, et al., 2015). Based on the current results, it is therefore not possible to know whether the results of the current studies generalize to any other stimulus than the video stimulus we used, or to social activities that do not include watching a video. Additionally, the results of the current study do not give insight into how the effect of sharing an experience compares to other factors influencing social closeness, such as for example eye-contact, which has to been shown to play an important role in chimpanzee interactions (Tomonaga, 2006; Tomonaga et al., 2004).

Furthermore, it is hard to tell if the apes’ psychological experience of watching a video together with a partner in the current study is cognitively similar to the psychological effects that occur when humans share an experience. Also, we do not
know if the short term psychological effects found in the current paradigm are sufficient
to influence great ape social relationships in the long term. Additionally, although the
approach and proximity measures tell us something about a general social interest in-
and/or motivation to interact with a social partner, the question remains how this
compares to the social closeness that humans feel after sharing an experience.
Nevertheless, the current results demonstrate that on a basic level, socially relating to
others via shared experiences seems not to be uniquely human but instead deeply
rooted in our evolutionary history.
Chapter 4. Human children, but not great apes become socially closer by sharing an experience in common ground\textsuperscript{13}

\textit{Introduction}

Most primates form social relationships with others through social activities that involve some form of physical interaction, such as grooming or physical play (Behncke, 2015; Hare et al., 2015; Lehmann et al., 2007). Humans, in addition, form social relationships with others psychologically, via social activities that create shared experiences. For example, humans have been shown to feel socially closer to others with whom they make music together (Pearce et al., 2015), dance together (Tarr et al., 2016), play team sports (Artinger et al., 2006), and converse, especially by sharing personal information (Aron et al., 1997) or attitudes (Bosson et al., 2006), or by gossiping (Dunbar, 2004). As these kinds of social activities are ubiquitous in humans, while other primates seemingly do not create social activities centered around sharing experiences solely for the purpose of creating social closeness, the question remains if sharing experiences works differently and/or has different consequences for humans than for other primates.

In humans, the effect of shared experience on social closeness to partners is so strong that it even occurs in minimally interactive shared experiences such as jointly attending to a video together (Wolf et al., 2015; although is effect is seemingly

\textsuperscript{13} This chapter is published in the \textit{Journal of Experimental Child Psychology} (Wolf & Tomasello, 2020b)
moderated by the content of the stimulus; Haj-Mohamadi et al., 2018; Rennung & Göritz, 2015). Recently, a similar effect of shared experience on social closeness has been found in young children (Wolf & Tomasello, 2020a). Two-and-a-half-year-old children watched a video on a screen. In one condition, an adult sat next to the child watching with them. In a control condition, the adult sat in the same location but did not watch the video: their view of the screen was blocked by a cardboard divider and, in any case, they were looking down reading a book. Afterwards, the adult sat across the room from the child playing with toys. The children who had viewed the video with the adult, approached the adult more quickly than those who had viewed the video by themselves while sitting next to the reading adult. This study thus demonstrates that the propensity to feel closer to those with whom one has shared an experience emerges early in human ontogeny.

In non-human primates there is also a relationship between coordinated behavior, communication and social bonds. There is no doubt that facial expressions signaling hostile or friendly intentions shape primate social networks. Additionally, multiple studies have reported that coordinated behavior, such as resting or sleeping in proximity, group travel and mobbing behavior are shaped by pre-existing social relationships within primate groups (Micheletta et al., 2012; Schamberg et al., 2016). Yet, there no evidence that such coordinated activities in themselves create social closeness. Additionally, there are no reports of non-human primates creating shared social
activities solely for the purpose of connecting with others in the way that humans do. It has therefore been suggested that, perhaps, human shared experiences allow us to socially connect with one another in unique ways (Dunbar, 2012; Wolf, Launay, et al., 2015).

Surprisingly, however, recent research using an experimental design similar to that of the shared experiences studies in humans found a similar effect in great apes (Wolf & Tomasello, 2019). In one experiment, chimpanzees and bonobos either watched a video with a human or watched the video by themselves while the human read a book in the exact same location (and was blocked from seeing the video). Replicating the results with children, the apes were faster to approach the human with whom they had watched the video in a subsequent interaction. In a second experiment with conspecifics, chimpanzees either watched the same video together or they sat in the same places but watched videos on different screens. Individuals spent more time in the same part of the room after having watched a video together than when they had watched videos on different screens. These findings potentially undermine the claim that the way humans create social closeness through shared experiences is a unique to our species.

In developmental psychology theorists have argued, however, that, human sharing involves some degree of recursive mindreading (Grice, 1957). It has been proposed that, in addition to just experiencing the same thing, a key part of humans sharing experiences is that participants know that they both know (i.e. have common
ground about the fact) that they are experiencing something together. Thus, while visually co-attending to something individuals establish common ground by, for example, a mutual look face-to-face, allowing both individuals to infer that they both are aware that they are attending to the same thing – and that the other is aware of this also (Bakeman & Adamson, 1984; Carpenter et al., 1998). That is, a gaze alternation in which individuals alternate looking at a stimulus with a mutual look allows one to infer that (1) the other is also attending to this stimulus, and (2) the other has seen one looking at them and at the screen, so that they know that one knows they have seen one looking at the screen. As such, one now knows that both individuals know that (1) they are looking at the same thing and (2) that they both know they both know they are looking at the same screen, knowledge that is often referred to as the common ground component of joint attention (Siposova & Carpenter, 2019; Tomasello, 2014, 2019). This common ground understanding is crucially important in many uniquely human cooperative activities, as it undergirds uniquely human forms of coordination in collaboration and shared reference in cooperative communication (Shteynberg et al., 2020; Tomasello, 2010a, 2019).

Our question in the current study was thus whether this common ground component plays a role in the creation of social closeness through shared experiences, and whether this is true for both human children and great apes alike. Given children’s propensity to create common ground with others seemingly for its own sake in a way
that apes do not do - for example, in pointing things out to others via a pointing gesture just to share attention to it (Liszkowski et al., 2004) - we expected children, but not apes, to be sensitive to a mutual look during the video watching. This is because the mutual look has the potential, for those who understand its significance, to create a common ground understanding that “we” are watching the video together.

To test this hypothesis, we conducted a comparative study in which participants watched a video at the same time as an adult human partner, while keeping the set up and procedure as comparable as possible between species. Participants entered a room where an experimenter was waiting for a video to start. In both cases, the participant was sitting behind the experimenter, with a physical barrier placed in between them. After the video the experimenter and participant separated for 15 seconds, after which they reconvened in the same room to engage in the dependent measure. Crucially, in the sharing condition the experimenter attempted to create common ground (i.e. that they were watching the video together) by turning and looking directly at the participant in response to the video starting. In contrast, in the observer (i.e. control) condition the experimenter did not turn during the video to look at the participant but instead looked at the participant the moment they reconvened in the same room. In this control condition, then, it was clear to the participant that each of them was attending to the same video, but the participant had no way of knowing whether the experimenter was aware of this as well; there was no information that allowed the participant to establish
common ground about the fact that they were watching the video together. As in previous research (Wolf & Tomasello, 2019, 2020a), we measured participants’ willingness to approach and engage in social behavior towards the experimenter in terms of approach latency after the manipulation. We predicted that children, but not apes, would approach faster in the sharing condition than in the observing condition.

**Study 1 – Children**

**Methods**

**Participants and design.**

The child study was approved by the universities Institutional Review Board (IRB). As the pilots of previous studies using an identical procedure for the dependent measure (Wolf & Tomasello, 2020a) had shown learning effect in children for this measure, we conducted an experiment with a 2 factor between subjects (sharing vs. observing) design. Based on our previous studies using similar measures we collected data until we had 64 (non-excluded) participants with gender counterbalanced across conditions.

In total 106 children from mostly middle-class Caucasian families living in a middle-sized US city participated ($M_{age} = 32.72$ months, $SD_{age} = 2.01$, 48 female). Two participants were excluded due to experimenter errors (e.g. E2 interacted with the participant when they were not supposed to), and four were excluded due to parent errors (e.g. the parent significantly deviating from to the script during the dependent
measure). In addition, 18 participants were excluded for child errors (e.g. fussiness), seven participants were excluded because they attend to the video for less than 30 seconds (i.e. 50% of the video), and nine participants were excluded because they approached neither in the baseline nor in the experimental approach measure. Finally, two participants had initially accidently been marked as excluded (i.e. a coding error) so that the data of two replacement participant was collected before discovering this error. To stay true to our original stopping time criterion we decided to use the two initial participants in the analysis of our counterbalanced sample of 64 participants. As such, the final sample consisted of 64 children between 2.5 and 3 years of age ($M_{age} = 32.65$ months, $SD_{age} = 1.96$, 32 female).

**Procedure.**

After the informed consent, E1 conducted a baseline approach measure to get an idea of a child’s propensity to approach a novel adult irrespective of the manipulation. The parent sat down on the floor on a mark, and the experimenter sat down 2 meters away while putting a marble run in front of them. E1 then offered a ball for the marble run to the child and said “if you want, you can go play!” We then measured how long it

In addition to the analysis reported in the results, we also ran the same analysis with a larger group, based on slightly more lenient exclusion criteria, which, in addition to the original 64 participants now also included children who had watched more than 15 seconds but less than 30 seconds of the movie ($N = 6$), children who waved at the experimenter during the baseline phase (but to whom E2 did not wave back; $N = 2$), and the 2 back-up participants who’s data were collected after their respective cells were already filled. The results for this group ($N = 74$), was not meaningfully different than the results for the original counterbalanced design of 64 participants (Estimated mean group difference of the larger sample: Bayesian estimation of Mean group difference: $Mode = 17.93$ seconds, $95\%$ HDI = 4.88 to 30.98).
took for the participant to approach the experimenter and take the ball from them. If the child did not approach after 90 seconds, we stopped the baseline measure and coded the time as 90 seconds (the maximum). If the participant did not approach, E1 took the marble run closer to the child, so that all children took the ball from E1 and played with the marble run before proceeding to the experimental phase. Shortly after the child had taken the ball from E1, E2 briefly entered the room to give the participant some exposure to E2 before the experimental trial. However, E2 did not interact in any way with the participant. Instead, E2 came in with a form, handed the form to E1 and said “Here is the form you wanted”. E2 then turned around and walked out of the room.

Next, E1 took the participant and the parent to the experimental room. Here, E2 was already sitting on a toddler chair on one side of the room, with their back to the door through which the participant entered, facing a TV screen on which a moving placeholder image was playing. At this point, the experimenter was not looking at the screen, but was instead reading their own book. The child sat down on a toddler chair behind E2, 1.5 meters from E1, at an angle at which they could easily see E2’s face to infer where E2 was looking, but where it was unlikely that E2 could see them unless they turned around. Furthermore, there was a row of toddler tables in between E2 and the participant so that the participant could not easily walk into the visual field of E2. Finally, the parent sat down on a pillow behind the participant to provide comfort in
terms of physical proximity, but was instructed to fill out some forms during the manipulation and not to engage with the child in the watching of the video.

Once the participant and parent had sat down, E1 started the video. The video was a one-minute video similar to a video used in previous research (Wolf & Tomasello, 2020a) containing short clips of animal behavior without sound. Crucially, in the sharing condition, when the video was started E2 turned around and made eye contact with the participant to establish common ground about the video having started. In the observer condition, E2 simply started looking at the screen.

After the video, E1 took the participant and the parent out of the room to “go look for some toys” until after 15 seconds of looking around outside the door E1 ‘realized’ that the toys were actually back in the room they just left. This part of the procedure provided a natural opportunity to incorporate an instance of eye contact between E2 and the participant in the observer condition that was not related to the watching of the video, so that the absolute amount of social engagement would be equalized across conditions. As such, when E1, the parent and the participant came back in the room, E2, who at that moment was busy putting away the pillow and toddler chair to make space for the dependent measure, stopped what they were doing for a second, looked towards the door and made eye contact with the child (as opposed to doing exactly the same without looking to the door in the shared condition).
Next, the participant sat down on a mark on the floor, flanked by the parent and E1 who was holding a stopwatch to keep track of the time during the dependent measure. E2 sat down on a mark 2 meters away from the participant, similar to what E1 had done during the baseline measure. Next, the participant engaged in the same approach latency procedure used in earlier studies (Wolf & Tomasello, 2020a). That is, E1 gave E2 a cue and started a stopwatch. E2 responded by taking out a stuffed animal from under a blanket and offering it to the participant. The moment the toy was offered E1 told the participant “If you want, you can go play with the toy...”, which the parent repeated after E1 following the instructions received before the warmup phase. If the child had not approached after 15 seconds E1 gave E2 another cue to give the stuffed animal a hug, and offer it again. After another 15 seconds, E1 gave a cue to E2, who responded by putting away the stuffed animal. E2 then took a toy truck with blocks from underneath a blanket and offered one of the blocks to the participant. Again, E1 cued E2 after 15 seconds. In this case, E2 tilted the back of the truck so that the blocks would fall out. E2 then grabbed another block and offered it to the participant. Finally, if the participant had still not approached, E1 cued E2 to put the truck away and take a marble run identical to the one in the warmup phase. E2 put the marble run next to them and offered a ball to the participant. During this part of the experiment, E2 looked at the participant, but never said anything to them.
Once the child approached (i.e. touched the toy E2 was holding) or if they had not approached after 90 seconds, the experiment was stopped. If the child did not approach within 90 seconds in both the baseline measure and the experimental measure, they were most likely too shy to engage in this procedure and therefore excluded from data-analysis. For a schematic overview of the child study, see Figure 8 (upper box).

---

Figure 8: Schematic overview of the manipulation for children (upper box) and great apes (lower box).

---

In addition to using the baseline measure as an exclusion criterion, we also used it to conduct the analyses on the same sample while controlling for individual differences in shyness. We first regressed the approach latency of the baseline phase on the dependent measure, after which we ran the analysis with the (Unstandardized) residuals. The results were not meaningfully different from our original analysis Bayesian estimation of Mean group difference: Mode = 17.76 seconds, 95% HDI = 5.13 to 30.39).
Coding

Every session was recorded by two cameras. Independent research assistants (i.e. research assistants who were not part of the procedure and thus unaware of what the participant did in the dependent measure) coded the baseline phase and the manipulation phase for exclusion criteria, such as experimenter errors, parent errors (e.g. the parent significantly deviating from to the script during the dependent measure), and child errors. Every trial marked for exclusion was checked by a second independent coder for confirmation. If the two coders did not agree, a third independent coder examined the trial and casted a final, deciding vote. Next, a different research assistant coded the approach latency, with a second research assistant coding 25% of the trials for reliability assessment (Intraclass Correlation Coefficient = .995).

Results

As the skewness and kurtosis of all dependent variables were between -1.96 and 1.96, we deemed the data normal enough for using Bayesian parameter estimation models with normal likelihoods and priors. A Bayesian independent samples test of group means (N = 64) showed an effect of condition on approach latency (Estimated mean group difference: mode = 16.41 seconds, 95% HDI = 2.49 to 30.33). Children in the sharing condition approached the experimenter faster (M = 16.70 seconds, SD = 21.00) than children in the observer condition (M = 33.11 seconds, SD = 32.49). See Figure 9 for the distribution of approach latency for the child sample.
**Study 2 – Great apes**

**Methods**

**Participants and design**

The ape study was approved by the local Animal Research Committee, and was done in accordance with all of the governing local laws and regulations concerning research with animal. The animals were housed in a zoo, where all groups spend different parts of their day in their sleeping rooms, a heated, tropical indoor enclosure, and a large outdoor enclosure which they can access when weather permits. The animals are regularly exposed to enrichment activities and all of them had previously engaged in social and cognitive experiments with humans. No food or water restrictions were associated with this study. Furthermore, for this study the maximum amount of diluted grape juice that could be offered to individuals during the study was established in cooperation with the zoo staff in charge of the animals’ diets.

Due to the limited availability of great apes, as well as a lack of learning effects in previous studies using the current dependent measure (Wolf & Tomasello, 2019), we conducted an experiment with a 2 factor within subjects (sharing vs. observing) design, where the participants’ trials were at least a week apart. In total, we tested 20 great apes, each of which engaged in both conditions. One chimpanzee did not approach in either condition and was therefore excluded from analysis. The final sample therefore consisted of 14 chimpanzees ($M_{age} = 28.00, SD_{age} = 10.64, 10$ female) and $5$ bonobos ($M_{age} =
17.20, $SD_{age} = 8.93$, 2 female). For additional information about the apes, see: Appendix B, Table 10.

**Procedure**

The procedure of the ape study was as similar as possible to that of the child study. However, due to practical and safety considerations, the procedure was different for great apes in some respects. The experiment was conducted in the sleeping rooms of the bonobos and chimpanzees, which have multiple segments separated by hydraulic doors. When a participant had entered the segment adjacent to the segment they would be watching the video in, E2 went into the room, allowing a brief moment of exposure for the participant, and sat down in front of a screen in a way similar to in the child study. Next, the keeper opened the hydraulic door (after which they left the room), at which point the participant entered the segment behind E2. Importantly, in each of the two trials the participant interacted with a different female experimenter whom they had never seen or interacted with before. E1 then offered the participant dilated grape juice through a juice tube with a mouth piece at the end (commonly used to motivate great apes to stay in place when doing studies involving video material, e.g. Krupenye et al., 2016; Wolf & Tomasello, 2019), which was available to the participant through a hole in the Plexiglas panel behind E2.

Once the participant had sat down, E1 walked out of the room to the laptop in the hallway that was connected to the screen inside the room and started one of several
minute silent videos of playing juvenile chimpanzees, similar to the videos previously used in similar studies (Wolf & Tomasello, 2019). There were different videos for different trials, so that all participants watched each video only once. To make sure the videos were highly similar in content, they were short 1 minute clips taken from a longer video in which showing a playing a juvenile chimpanzee with other juvenile chimpanzees and an adult chimpanzee sitting in the background. The manipulation in the sharing condition was identical to that of the child study: when the video started E2 turned around and looked back at the participant, making brief eye contact to allow the participant to infer that they both know they were both watching the video.

Next, just like in the child study, we provided an opportunity for E2 to make eye contact with the participant in the observer condition. That is, after the video was finished, E2 got up and briefly left the room and came back after 15 seconds. In the sharing condition, E2 walked back in without looking at the participant and sat down at the other end of the room. In the observer condition, when coming in E2 socially engaged the participant by means of a communicative look, looked away again, and then proceeded to walk on to the side of the room where they would be sitting during the dependent measure. In both cases, E2 was holding a stopwatch to keep track of the time during the dependent measure.

Next, like in previous studies (Wolf & Tomasello, 2019) and the current child study, we measured the participants’ willingness to approach E2. Similarly to previous
studies using this measure with great apes, approach latency was operationalized as the moment the participant entered the area four tiles behind the Plexiglas screen behind which E2 was now sitting (similar to previous studies, see: Wolf & Tomasello, 2019). If the participant did not approach after 30 seconds, the experimenter started brushing their fingers along the mesh to prompt the participant to come over. If they still had not approached after 60 seconds, the experimenter, in addition to rattling the mesh, also called out their name. The moment the participant entered the approach area, or if the participant did not approach within 90 seconds, the trial was finished. If participants did not approach in one of the two conditions, their time was coded as the maximum time (i.e. 90 seconds). If a participant did not approach in either trial, they were excluded from data analysis. For a schematic overview of the ape study, see Figure 8 (lower box).

Coding

Three cameras recorded the sessions. The footage from a camera under the TV screen was used to capture (and check) the look between E2 and the participant. An overview camera recorded the overall movement of the experimenters, keeper and participant, and a ceiling camera positioned above the approach area was used to determine if the participant stepped into the approach area. One independent research assistant coded the approach latency for all individuals from the ceiling camera (from which it was impossible to discern which condition the participant was in), and a second reliability coder coded 25% of the trials (Intraclass Correlation Coefficient = .995).
Results

As the skewness and kurtosis of all dependent variables were between -1.96 and 1.96, we deemed the data normal enough for using Bayesian parameter estimation models with normal likelihoods and priors. A Bayesian independent samples test showed no effect of the order in which the conditions were presented on the difference in approach latency between conditions (Estimated within subjects difference: Mode = 12.98 seconds, 95% HDI = -32.84 to 58.81 seconds), suggesting that, like in previous research using the same measure with great apes in a within subjects design (Wolf & Tomasello, 2019) there were no order effects. Crucially, a Bayesian related samples test estimating the difference score between conditions did not show an effect of condition on approach latency (Estimated within-subjects difference: Mode = 3.01 seconds, 95% HDI = -19.52 to 25.55 seconds). That is, the apes approached the experimenter roughly equally fast after the sharing condition ($M = 38.02$ seconds, $SD = 32.15$) and after the observer condition ($M = 35.01$ seconds, $SD = 33.76$). See figure 9 for an overview of the distribution of approach latency of the apes.

Additional analyses showed no difference between chimpanzees ($N = 14$) and bonobos ($N = 5$) in their between condition difference scores (Estimated within subjects difference: Mode = -9.54 seconds, 95% HDI = -54.8 to 34.7 seconds), nor between participants that were hand reared ($N = 5$) and those who had been reared by their
parents (N = 12)\textsuperscript{16} (Estimated within subjects difference: Mode = -18.4 seconds, 95\% HDI = -77.8 to 38.0 seconds). A cautious interpretation of these between-subjects comparisons is, however, warranted, given the unbalanced distribution of individuals across the groups within these comparisons.

\textbf{Figure 9:} Distributions of approach latency per condition for children (left) and great apes (right).

\textsuperscript{16} The rearing history of 2 individuals was unknown
Discussion

The current results highlight a subtle but important difference in the way that human children and great apes experience situations in which they are co-orienting with a partner to a visual stimulus. Great apes experience it the same way whether or not there is a confirming look from the partner, with eye contact, whereas children experience these two situations differently, potentially suggesting that humans share experience in a way that other apes do not.

To interpret these findings, most notably the absence of an effect for the great apes, it is important to evaluate whether apes are in principle capable of correctly inferring the attentiveness of the experimenter towards the video and towards themselves during the manipulation. Although initial research questioned ape’s capacity to do so (Povinelli & Eddy, 1996a, 1996b), subsequent work has shown that these results were most likely a methodological artifact (Kaminski et al., 2004), and that apes do in fact understand what others can and cannot see (Bräuer et al., 2007; Hare et al., 2000, 2001), and do so in a context in which the individual’s line of sight goes through a transparent window as opposed to an opaque barrier (Okamoto-Barth et al., 2007; Tempelmann et al., 2011).

As such, it seems that apes should in principle be able to correctly infer that (1) the experimenter attended to the screen and (2) the experimenter attended to them. A related concern might be that the dependent measure itself is not appropriate for great
apes (but only for human children). However, previous studies (conducted in the exact same physical environment as the current study) have shown that both children and great apes are, in principle, sensitive to this approach latency based dependent measure (Wolf & Tomasello, 2019, 2020a). More specifically, these studies show that the current dependent measure is sensitive specifically to an experimental procedure in which the attention of an experimenter is manipulated, making it particularly appropriate for the current study.

Of course, as always, methodological differences between the child and ape studies could be at play. The most obvious one in this case, that children are dealing with a conspecific and apes are dealing with a human, is a concern. A conspecific partner was not possible in the current study because the partner’s behavior was carefully prescribed. Yet, in a previous highly similar study showing that great apes show more social closeness after co-orienting to the same stimulus (Wolf & Tomasello, 2019), the effect was the same for both a human and a conspecific partner.

Our conclusion is thus that the most likely explanation for our current results is that the children, but not the apes interpreted the look differently between conditions. That is, the children, but not the apes understood that the experimenter’s look to them in response to the onset of the video, followed by a look back to the video was an attempt to constitute a common ground understanding that they were watching the video together. In contrast, when the experimenter, in the control condition, looked to the
child they interpreted it as merely a response to noticing another individual in the same room when the experimenter and participant reconvened. For children to understand the two looks differently – such that it had a significant effect on their willingness to interact with the experimenter – required them to understand that mutual looks in response to a stimulus onset allows individuals to infer that both individuals now know that they both know that their attention was drawn by this same stimulus, thus allowing individuals to create common ground about the onset of this stimulus.

In order to understand the difference between this, and a situation in which individuals look to one another because they merely are salient stimuli to one another (i.e. after reconvening in the same room) children thus needed to align and coordinate mental states with the experimenter cognitively (perhaps recursively) in ways that, apparently, were not available to the great apes. That is, the creation of common ground with a partner might require a unique form of recursive mental coordination or some other way of establishing socially shared experiences, perhaps through creating a collective agent (see: Shteynberg et al., 2020), although the question remains if such a collective agent can be construed without the capacity process recursively structured social information.

The current results might therefore help to explain why humans, but not other species go to such great lengths to create and engage in social activities centered around shared experiences. Previous research showed that on a basic level simply co-attending
to something creates social closeness in both humans and great apes. However, the current results show that, in addition to this co-attending mechanism, humans, but not great apes, create more social closeness when they create common ground about their experience being shared than when they are merely observing someone co-attending to the same thing. This suggests that the human motivation to spend large amounts of time, energy and resources in shared social activities might not just be driven by a desire to go through similar experiences with others, but instead by the unique opportunity such activities provide for creating shared common ground with others, which serves both to create social closeness and to support many further cooperative and cultural activities.

More generally, the current results provide support for a uniquely human form of shared cognition which in this particular case, is operating in the realm of social closeness. It would seem that only humans have evolved the psychological processes (or at least sensitivity) necessary for creating common ground about things we experience together. As this common ground plays a crucial part in a variety of human social dynamics such as sophisticated (e.g. role based) behavior coordination, collaboration, and cooperative communication, the current results help explain how and why human social cognition and behavior is different from that of other primates.
Chapter 5. Creating social closeness in adult dyads and groups through joint attention in video-mediated social interactions.

Introduction

Humans are remarkably effective at maintaining large social networks (Dunbar, 1991; Dunbar et al., 2009; Lehmann et al., 2007). This is said to be, at least in part, due to humans’ unique capacity to engage in shared cognition, which allows us to participate in social activities centered around shared experiences in order to create social closeness with others (Wolf, Launay, et al., 2015; Wolf & Tomasello, 2019, 2020a, 2020b). In recent human history, however, technological advancement has dramatically changed the way in which we interact with others, most notably in the context of digital communication. This novel social context is forcing our social cognitive machinery, which adapted to accommodate our in-person interactions, to operate in a context in which detailed information about our interaction partners’ behavior and/or environment is not as readily available to us. The question is, therefore, whether human social cognition is flexible enough to accommodate this change. Do we still effectively create social closeness in interactions mediated by digital technology? And, if so, is this social bonding mechanism in this setting still effective when attempting to create social closeness within a group?

Creating and maintaining social relationships is crucial for humans to maintain good mental and physical health (Baumeister & Leary, 1995). Compared to other
animals, humans seem to be particularly effective at maintaining large, cooperative social networks (Dunbar, 1991; Dunbar et al., 2009; Lehmann et al., 2007). One reason for this is that, compared to for example other primates, humans seem to have additional, seemingly unique social bonding mechanisms that allow them to create social closeness with others. Aside from creating social closeness in conversations in which we disclose personal information (Aron et al., 1997) or gossip about others (Dunbar, 2004), we connect with others by singing together (Pearce et al., 2015; Weinstein et al., 2016), dancing together in synchrony (Cirelli et al., 2017; Hove & Risen, 2009; Tarr et al., 2015, 2016), playing team sports together (Artinger et al., 2006) and participating in rituals together (Charles et al., 2021; Singh et al., 2020).

Over the last decades, scholars have been trying to find psychological explanations for why (seemingly) only humans engage in such peculiar social bonding activities. Recently, researchers have raised the possibility that one of the core mechanisms underlying our capacity to connect with others through shared social activities is shared cognition: a set of social cognitive mechanisms underlying humans’ capacity to make sophisticated inferences about others’ mental states in relation to their own (Tomasello, 2014, 2019; Tomasello & Carpenter, 2007).

This idea is supported by research examining social bonding in minimally interactive shared experiences through joint attention. Studies have shown that human adults who attend to something together feel socially closer to their partner than
towards a partner who was attending to something else (Haj-Mohamadi et al., 2018; Rennung & Göritz, 2015; Wolf, Launay, et al., 2015). Follow-up research showed a similar effect in young children (Wolf & Tomasello, 2020a) and, surprisingly, also in great apes (Wolf & Tomasello, 2019), although an additional comparative study showed that this works somewhat differently for human children than for great apes (Wolf & Tomasello, 2020b).

Overall, these findings suggest that, indeed, humans have evolved sophisticated social cognitive mechanisms that allow them to create social closeness through social interactions in shared social activities in unique ways. However, over the last couple of decades, novel means of social interaction have rapidly become available to us. Most notably, the introduction of digital and online communication has profoundly shaped the way we are able to interact with others. The question is, therefore, whether the shared cognition that has facilitated social bonding in face to face interactions and in-person social activities is flexible enough to accommodate social bonding in these novel, digital social contexts.

One example that illustrates the potential problems our social cognition is faced with is the fact that in online communication, even in video chat, making eye contact is impossible. After all, this would require the camera to be positioned inside the part of the screen that displays the partner’s eyes. As eye contact has been shown to be important for both adults and children to judge the degree of closeness and
cooperativeness of a partner (Cui et al., 2019; Kleinke, 1986; Siposova et al., 2018; Wolf & Tomasello, 2020b), the impossibility of making eye contact in these online settings might make it more difficult to connect to others in these types of interactions.

There is, however, some evidence that humans are able to create social closeness through computer mediated interactions. One on one interactions online or through text message can create trust between participants (Zheng et al., 2002) or, conversely, elicit feelings of ostracism (Smith & Williams, 2004). Additionally, online gaming with partners causes individuals to evaluate these partners more positively (Dabbish, 2008), but can also cause feelings of social exclusion (K. D. Williams et al., 2000; K. D. Williams & Jarvis, 2006). Furthermore, recent research on online gaming corroborated these findings, showing that online multiplayer games can increase social closeness between participants (Depping & Mandryk, 2017). Specifically, this study showed that cooperation (i.e., behavior coordination) and feelings of interdependence are driving factors behind the creation of social closeness in online games. This raises the question if online social bonding is constrained to social settings characterized by cooperation and interdependence, or if inferring that an experience is shared, for example through joint attention, is enough to create social bonding in an online social interaction, as is the case in in-person social interactions (Haj-Mohamadi et al., 2018; Rennung & Göritz, 2015; Wolf, Launay, et al., 2015; Wolf & Tomasello, 2019, 2020a, 2020b)?
Moreover, if this mechanism does work in online social interactions, does this still work in the context of a larger group? Human social activities such as making music together or playing games together have been argued to be an efficient social bonding mechanism because they allow participants to create social closeness with multiple individuals at the same time (Dunbar, 1998, 2004, 2008, 2012). One study directly demonstrated this by showing that even in very large groups (i.e., 232 individuals) singing together still created social closeness between group members (Weinstein et al., 2016).

To find out whether joint attention creates social closeness in online social interactions, and, if so, whether this social bonding mechanism in this context is also effective in a group setting, we conducted an experiment in which participants engaged in an online video-mediated social interaction, either in a dyad or in a group of four. In addition, we also manipulated whether participants engaged in a joint or disjoint attention experience, a manipulation which was designed to be as similar as possible to previous studies conducted in-person. In the joint attention condition participants engaged in a minimally interactive shared experience, jointly attending to a video on a shared screen. In the disjoint attention condition, participants were told that they would watch the video sequentially, and that the other participants (which were pre-recorded confederates) would watch the video after they had finished. After the manipulation, we measured participants’ self-reported social closeness to the other participants.
Methods\textsuperscript{17}

Participants and design

Participants between 18 and 30 years old were recruited online through a variety of databases that included students and alumni of the local university, as well as members of the local community. We aimed for 24 included participants per cell in our 2 (Attention: Joint vs Disjoint)\textsuperscript{18} by 2 (Group size: Dyad vs Group) design, meaning 96 participants in total. We collected data from 139 participants, 43 of which were excluded for a variety of reasons, such as inattentiveness (i.e., failing the comprehension questions or instructional manipulation check, not internalizing the manipulation or indicating to be inattentive during the manipulation; \(n = 28\)), not following instructions (\(n = 6\)), indicating that they were suspicious that the other participants were confederates (\(n = 6\)), technical difficulties (\(n = 2\)), and experimenter error (\(n = 1\)). As such, our final sample, as planned, included 96 participants (\(M_{\text{age}} = 19.72, SD_{\text{age}} = 1.55\)), 66 of which identified as female, 29 as male and 1 as non-binary. Furthermore, participants’ ethnic backgrounds included Asian (\(n = 44\)), White/Caucasian (\(n = 34\)), Mixed (\(n = 7\)), African American (\(n = 6\)), and Hispanic (\(n = 4\)), while 1 participant did not answer this question. Participants were rewarded with a $10 gift card or course credits.

\textsuperscript{17} This study was pre-registered at: https://osf.io/p85wh/?view_only=f8288c523fd4d4076baf349d0bde22bf4
\textsuperscript{18} This part of the study procedure was an honors thesis project by Kayley Dotson
Procedure

After registering for the study, participants received an email in which they were told that they would be taking part in an online video call (i.e., through Zoom) with other participants. They were asked to use a laptop or desktop to join. In this email they also received a link to a survey that contained an informed consent form and questions about their demographics (i.e., age, gender, race, and student ID, if they had one). Furthermore, participants were told that all of the participants in the study would first meet with the experimenter separately to go over some instructions and to make sure everyone’s set up worked properly. Participants saw a list of names which included their own, with for each name a specific time indicated to join the meeting. In reality, participants always received an email with their name last on the list, so that the experiment always started immediately after the experimenter had gone over the instructions with them.

Once participants joined the meeting, they were welcomed by the experimenter who went through the instructions with them. To standardize the experience as much as possible for participants, they were asked to (1) have their video enabled and (2) their microphones muted once the study started, (3) only have one monitor active during the study, (4) make sure that their names were displayed on top of their video in Zoom, and (5) make sure that they could see all the participants in the call displayed on top of

In total 3 different experimenters ran the study. Adding experimenter in as a factor in the analyses showed that results did not meaningfully change as a function of specific experimenters.
the screen the experimenter was sharing. Furthermore, the experimenter also checked whether participants knew (1) how to use the chat feature in Zoom, (2) how to open a link from the chat feature of Zoom to display in a browser on their computer, and (3) how to make a screenshot of the Zoom call and save this screenshot for later use.

During this initial stage, participants also received instructions for during the study. They were told that after the instruction was over the experimenter would have all participants join the meeting, after which the experimenter would share a video about animals from the National Park Service through a shared screen. The experimenter also indicated that they would disable their own camera and microphone right before the video would start, to ensure that they would not disturb the experiment. In addition, the experimenter told the participants that if they wanted to contact them, they could send them a private message over Zoom chat.

Crucially, in the joint attention condition participants were told that all the participants in the Zoom call would be watching a video on a shared screen, whereas in the disjoint attention condition they were told that one person at a time would watch the video, sequentially, while the other participants could do something else while they waited, as long as they remained visible on their camera. In reality, the participant always went first (i.e., in both conditions the participant always watched the video at the start of the experiment). While giving these instructions to participants in the disjoint attention condition the experimenter suddenly ‘realized’ that the participant would be
watching the video first, and that some of these instructions were therefore not particularly relevant to them. In reality, this instruction was meant to make clear to the participant that the other participants might be doing something else while the participant was watching the video.

Once the experimenter had finished the instructions and asked the participants if they had any more questions, the experimenter let the ‘other participant(s)’ out of the breakout room. In reality, these other participants were pre-recorded videos of student aged individuals who were not part of the local student community (to decrease the risk of pre-existing social relationships between the participant and the confederates). Sufficient male and female confederate videos were created so that the gender of the confederates could be matched to the participant (similar to in previous in-person research, see: Wolf, Launay, et al., 2015). Crucially, in the dyads only one additional participant entered the call, whereas in the larger group, three additional participants entered the call. To create this group, the experimenter controlled either two (small group) or four (large group) computers, where all but one computer (i.e., the one the experimenter was using to interact with the participant) had an opaque piece of paper taped over the camera. This then allowed the experimenter to play the videos of the confederates as background videos in Zoom in a way that made the confederates appear to be real life participants (See Figure 10).
Once the other participants had joined, the experimenter briefly welcomed the participants and asked everyone if they could still hear them. To make the confederate videos more credible as real life participants, these two sentences were timed so that the videos of the confederates gave a thumbs up and/or a nod immediately after this question, making it seem like they responded to the experimenter’s question. Next, the experimenter started the video. In the joint attention condition, the confederates in the videos kept paying attention to their screen(s), whereas in the disjoint attention condition, the confederates disengaged from their screen(s) and started looking at their phone (see Figure 10).

Figure 10: Example of the participants’ perspective during the manipulation for each of the four cells in the design.
After the video had finished, the experimenter sent a private message to the participant containing the link to the survey and their participant number, which participants filled in on the first page of the survey. Next, participants were asked to make a screenshot of the Zoom call and save it so they could upload it later in the survey. They were then instructed to leave the Zoom meeting and completed the rest of the survey, which contained comprehension questions (i.e., attention manipulation checks), questions about the experience of watching the video, social bonding questions about the other participants, and several additional questions and manipulation checks. Finally, participants read the debriefing and exited the survey, after which they were compensated within 24 hours of their participation.

**Measures**

**Attentiveness manipulation check**

To catch inattentive participants, we asked them three questions about the content of the video, namely what the main subject of the video was (i.e., mountains, animals, airplanes or beaches), whether or not there was a monkey in the video (true), and whether or not there was koala bear in the video (false). Furthermore, we also asked participants how many participants (including themselves and the host) were in the Zoom call during the experiment (i.e., three, four, or five, with the correct answer being three for the dyads and five for the larger groups). Participants that answered these questions wrong were excluded from data analysis.
Watching experience

To get a sense of how participants experienced watching the video, we asked them to indicate on a 100-point slider scale how much they liked watching the video, how much they enjoyed watching the video, and how much they liked the video (0 = “not at all”, 100 = “a lot”). Furthermore, we also asked them how often they were distracted during the video (0 = “never”, 100 = “all the time”). Finally, we asked participants whether they would like to watch the video again by themselves and how attentive they were during the video (0 = “not at all”, 100 = “very much”).

Instructional manipulation check

In the midst of the questions on watching experience, we included an instructional manipulation check (Oppenheimer et al., 2009). Participants were presented with an item that said “Please place the slider on exactly ‘37’. This is used to spot inattentive participants”. Participants that did not follow these instructions were excluded from data analysis.

Social bonding

Participants were first asked to enter the name of (one of) the participant(s). They were encouraged to, if they had forgotten the name, look at the screenshot they took earlier. Participants who put down the name of the experimenter were excluded from the sample. Next, to measure social bonding, we asked participants to answer 8 questions on how they felt about this other participant (i.e., separate for each participant
in the large group condition) on a 100-point slider scale. These questions were almost\textsuperscript{20} identical to the questions of the social bonding scale of previous research (Wolf, Launay, et al., 2015). Specifically, we asked participants how much they liked the other participant and to what extent they thought the participant was liked by others (0 = “not at all”, 100 = “a lot”), how positively they felt towards that participant (0 = “very negatively”, 100 = “very positively”), how much they trust this participant and how much they connected with them (0 = “not at all”, 100 = “very much”), how cooperative they felt towards that participant (0 = “not at all”, 100 = “very cooperative”), to what extent they felt close with that participant (0 = “not at all close”, 100 = “extremely close”), and, if they had to do a similar task again, how they would feel about doing that task again with this participant (0 = “I would prefer to do it with someone else”, 100 = “I would prefer doing it with the same person”). These questions were then collapsed into a single social bonding score, and for the larger group cells, averaged across the three confederates. Analyses showed acceptable reliability for this scale (Posterior mean of Bayesian Cronbach’s alpha = .750, 95% HDI = .672, .823).

\textsuperscript{20} The only difference was that we replaced the Inclusion of Other in Self scale with a more generic question about perceived social closeness. This new question seemed more intuitive in a setting in which participants were not in close physical proximity, and was also more consistent with the rest of the scale in terms of response format.
Additional questions and manipulation checks.\textsuperscript{21}

At the end of the survey, we also asked participants how impolite they felt the other participant was (0 = “not at all”, 100 = “very impolite”), how attractive they found this participant (0 = “not at all”, 100 = “very attractive”), and how aware they were of the other participants during the video (0 = “not at all”, 100 = “very much”). Participants were then also asked about their sexual orientation (“heterosexual”, “Lesbian”, “Gay”, “Bisexual”, “Queer”, “Asexual”, “Other”, or “Prefer not to say”). Finally, at the very end of the survey, we asked participants two open ended questions, namely (1) “what were the other participants instructed to do during the video, and did they follow these instructions?”, and (2) “Sometimes people develop ideas about what studies are about and what researchers are trying to find. If you have any thoughts about what we are studying, please describe them below”. These questions were meant to see if participants were suspicious about the confederates in the experiment and/or were aware of what the manipulation was trying to achieve. Suspicious participants were excluded from data analysis.

\textit{Results}

The standardized residuals of the social bonding scores (skewness = .29, kurtosis .40) indicated that the data was appropriate for using standardized Bayesian models.

\textsuperscript{21} The items about impoliteness, attractiveness and sexual orientation were not part of the pre-registration, but added after the registration (but before the start of data collection) based on external feedback, as they were potentially important confounds to control for.
with uninformed priors and normal likelihoods in JASP version 0.14.1 (Marsman & Wagenmakers, 2017).

**Social bonding**

To estimate the effect of the attention manipulation and group size on social bonding scores we conducted a 2 x 2 between subjects Bayesian ANOVA. Results showed that of all the possible models, the Null model (i.e., without any predictor variables) was by far the most likely one, showing no main effect of the attention manipulation or group size, nor an interaction effect on social bonding scores (see Table 5). The parameter estimates of the mean social bonding scores show highly overlapping 95% Bayesian High Density Intervals (HDI’s) for the participants who engaged in a dyadic interaction in the joint attention condition ($M = 44.72, 95\% \text{HDI} = 40.95, 48.49$) and in the disjoint attention condition ($M = 45.05, 95\% \text{HDI} = 39.61, 50.49$), as well as for participants who engaged in a group interaction in the joint attention condition ($M = 47.59, 95\% \text{HDI} = 43.42, 51.77$) and in the disjoint attention condition ($M = 43.87, 95\% \text{HDI} = 38.99, 48.75$). See Figure 11 for a graphical representation of the 95% HDI’s per cell.
Table 5: Model Comparison parameters for social bonding models.\textsuperscript{22}

| Models                          | P(M) | P(M|data) | BF\textsubscript{M} | BF\textsubscript{10} | error % |
|--------------------------------|------|----------|---------------------|--------------------|--------|
| Null model                     | 0.200| 0.627    | 6.723               | 1.000              |        |
| Attention                      | 0.200| 0.174    | 0.845               | 0.278              | 0.037  |
| Group size                     | 0.200| 0.143    | 0.670               | 0.229              | 0.034  |
| Attention + Group size         | 0.200| 0.040    | 0.167               | 0.064              | 1.062  |
| Attention + Group size + Attention * Group size | 0.200| 0.015    | 0.062               | 0.024              | 2.252  |

\textsuperscript{22} We also ran models that included the gender of the participants and the different videos that were shown to the participants in the Dyads as additional factors, but this did not yield meaningfully different results.
Figure 11: Bayesian High Density Intervals containing the 95% most likely social bonding scores for the Joint Attention and Disjoint attention groups per group size.
To further corroborate the absence of the effect of the attention manipulation, we conducted Bayesian independent samples T-tests of equivalence\textsuperscript{23}. A Bayesian equivalence test comparing the difference between the joint attention and disjoint attention condition in the full sample\textsuperscript{24} showed very strong evidence for the two conditions being practically equivalent (BF = 48.06). We also conducted separate equivalence tests for the participants who interacted in a dyad and participants who interacted in a group. The equivalence test for the participants who engaged in a dyadic interaction showed strong support for the social bonding scores in the joint attention and disjoint attention condition being practically equivalent (BF = 26.58). In addition, the equivalence test for the participants who engaged in a group interaction showed moderate support for the two groups being practically equivalent (BF = 5.92). Overall, these findings suggest that the current attention manipulation had no effect on the participants’ social bonding scores.

\textsuperscript{23} In Bayesian equivalence T-tests, like in regular Bayesian T-tests, the posterior distribution of the difference score between two groups is estimated. However, instead of looking at the 5\% most extreme values of the difference score at the tails of the distribution to infer non-equivalence, equivalence tests look at the 5\% most extreme values centered around zero, creating a region of practical equivalence. This helps to distinguish null-effects caused by high uncertainty in parameter estimation (e.g., due to low power) from null-effects that actually reflect practical equivalence between two groups. This equivalence model (i.e., the mean difference falls within the region of practical equivalence) is then compared to a non-equivalence model (i.e., the mean difference falls outside of the region of practical equivalence) in terms of their relative likelihood, using a Bayesfactor, which indicates the odds ratio in favor of the equivalence model or the non-equivalence model.

\textsuperscript{24} Bayesian equivalence tests are a relatively new phenomenon, and there is therefore currently no method to compute equivalence tests in models with more than one independent variable.
**Additional control variables**

Additional Bayesian 2x2 ANOVA’s showed no main effects or interactions (i.e., the Null model was the most likely model) of the joint versus disjoint attention condition or the dyad versus group setting in the degree to which participants felt the confederates were behaving in an impolite manner (see Appendix C, Table 11), or the degree to which the participants felt the confederates were attractive (see Appendix C, Table 12). As the perceived impoliteness and attractiveness of the confederates was similar between conditions, it is unlikely that these factors would have meaningfully influenced the current results.

**Watching experience**

As the questions how much participants liked watching the video and how much they enjoyed watching the video were conceptually very similar and highly correlated (Bayesian estimation of correlation coefficient: .962, 95% HDI = .942, 974) we collapsed them into a single viewing enjoyment measure. A Bayesian 2x2 ANOVA showed no main effects or interaction effect of the Attention manipulation or group size on viewing enjoyment (see Appendix C, Table 13), nor on the degree to which participants liked the video (see Appendix C, Table 14), or how much the participants would like to watch a similar video by themselves (see Appendix C, Table 15).
Discussion

The current results do not show a social bonding effect of sharing attention through online video-mediated social interactions. Crucially, Bayesian equivalence tests suggest that this was not just a matter of insufficient power, but instead showed that, on average, there was strong support for the social bonding scores to be practically equivalent for participants who were in the joint attention condition and those who were in the disjoint attention condition. In a similar fashion, there was no effect of group size, nor did group size moderate the effect of the attention manipulation on social bonding scores.

These results are in stark contrast with the social bonding effect of joint attention found in studies using real life social interactions (Haj-Mohamadi et al., 2018; Rennung & Göritz, 2015; Wolf, Launay, et al., 2015; Wolf & Tomasello, 2019, 2020a, 2020b). The question is why the studies done in these different contexts (i.e., online vs in person) yield different results. Clearly, there are important differences between these in-person studies and the online study in the current work. The pivotal question is, therefore, whether social bonding through sharing attention in online video-mediated interactions simply does not work as well as it does in real life? Or did, perhaps, the research logistics involved with collecting data online negate the effect found in previous in-person studies in an artificial (i.e., non-ecologically valid) way? In order to make an argument for the former interpretation, it is important to establish what exactly the
differences between previous in person studies and the current online studies were, and, crucially, whether these differences are, in fact, ecologically valid differences inherent to the social dynamics of interacting with others online versus in person, or, instead, specific to the logistics and set-up of the online data collection in the current study.

The design and procedure of the current study was kept as similar as possible to previous studies examining the social bonding effect of joint attention in in-person social interactions. Previous studies on this phenomenon have successfully used both between- (Haj-Mohamadi et al., 2018; Rennung & Göritz, 2015; Wolf & Tomasello, 2020a, 2020b) and within- (Wolf, Launay, et al., 2015; Wolf & Tomasello, 2019) subjects designs. As such, it seems unlikely that the choice for a between-subjects design in this particular study will have meaningfully influenced the results. Furthermore, with regard to the content of the manipulation, using video stimuli to create joint attention is common practice in studies on the social bonding effect of joint attention (Dunbar et al., 2016; Haj-Mohamadi et al., 2018; Rennung & Göritz, 2015; Wolf & Tomasello, 2019, 2020a, 2020b), several of which used animal videos with content similar to the one used in the current study (Wolf & Tomasello, 2019, 2020a, 2020b). Moreover, the social bonding measure was almost identical to that of a previous study with adults on the social bonding effect of joint attention in an in-person social interaction context (Wolf, Launay, et al., 2015). Also, similar to in previous studies, the social bonding scale was administrated through 25

 Aside from a single question regarding social closeness. See footnote 21 in the methods section.
a Qualtrics survey and was complemented by questions about the participants’ experience of watching the video itself.

The most important aspect of the study that needs to be examined in the context of the current results, however, is the degree of similarity of the joint versus disjoint manipulation set-up between the current study and those of previous in-person studies. During the manipulation of the current study, participants could clearly see that their partners were either looking at their own monitor (i.e., watching the shared screen) or were not looking at their screen and were looking at something else (i.e., looking at their phone). As such, the nature of the difference between the joint and disjoint condition, through gaze direction, was very similar to previous in-person studies on the social bonding effect of joint attention (Haj-Mohamadi et al., 2018; Rennung & Göritz, 2015; Wolf, Launay, et al., 2015; Wolf & Tomasello, 2019, 2020a, 2020b). As such, it seems unlikely that the lack of difference between the joint and disjoint attention condition in the current online study can be attributed to participants in the disjoint attention condition of the current study having a harder time inferring whether or not their partners were looking at their distractor (i.e., their phone) than in the disjoint attention conditions of previous studies.

There are, however, some small differences in set-up between the previous in-person studies on the social bonding effect of joint attention and the current online study that might, at least in part, explain their divergent results. Participants in the video-
mediated paradigm could, for example, not see the screen that their partners were seeing. Instead, they could only assume that their partners were following the instructions of the experimenter to not make the Zoom interface smaller while watching the video on the shared screen in order to do something unrelated to the experiment. In contrast, participants in previous in-person studies were watching stimuli on the same screen as their partner (Dunbar et al., 2016; Haj-Mohamadi et al., 2018; Rennung & Göritz, 2015; Wolf, Launay, et al., 2015; Wolf & Tomasello, 2019, 2020a, 2020b), making it relatively easy to infer what exactly their partner was seeing. This added uncertainty for participants in the current study about what their partner was seeing during the manipulation might have made it more difficult for them to experience joint attention the same way participants in previous in-person studies did.

In addition, as the video stream of the conversation partners in the current study was situated very close to (i.e., right above) the part of the screen on which the content of the shared screen was displayed, participants in the current study might have had a hard time distinguishing when their partners were looking at them and when their partners were looking at the video. In contrast, this was much easier for participants in

---

26 In theory, this would have been even more difficult for participants in the group condition. In the individual condition, participants knew that they were the only other person in the conversation whose camera was turned on. Thus, if their partner looked to the area above the shared screen, they were most likely looking at them. In the group condition, however, there were two additional participants present in the conversation, making it even harder to determine whether a partner was looking at them. Despite this added difficulty the results for the Group condition were not different from that of the Dyad condition, suggesting that this added difficulty did not meaningfully change the results.
previous in-person studies on the social bonding effect of joint attention. In these studies, the source of the stimuli to which partners jointly attended was always placed in a location clearly distinguishable from the location of the participants and their partners, so that it was relatively simple for participants to infer if and when their partners were looking at them or the stimulus (Dunbar et al., 2016; Haj-Mohamadi et al., 2018; Rennung & Göritz, 2015; Wolf, Launay, et al., 2015; Wolf & Tomasello, 2019, 2020a, 2020b). As such, relative to previous in-person studies, it might have been more difficult for participants in the current study to infer whether their partners were actively processing that they were watching the video together (or not). This uncertainty might have, at least to some degree, impaired the creation of social closeness between participants and their partner in the current study.

A related, albeit slightly different issue is that in contrast to participants in previous in-person studies on the social bonding effect of joint attention, it was impossible for participants in the current study to make eye contact with their partners, simply because in typical video communication set-ups the camera is not in the same place as the image of interaction partners’ eyes. Although eye contact in triadic interaction is not a necessary prerequisite for creating social closeness per se (Wolf, Launay, et al., 2015; Wolf & Tomasello, 2019, 2020a), it does play an important role in inferring the degree of social closeness between individuals (Cui et al., 2019; Kleinke, 1986; Siposova et al., 2018) and facilitates the creation of social closeness independently
of inferences of attention convergence in the context of shared experiences, at least in humans (Wolf & Tomasello, 2020b). As such, not being able to capitalize on this social dynamic might have, at least to some degree, contributed to the attention manipulation in the current study being less effective.

Finally, participants in the current study had considerably less information about the peripheral environment of their partner than participants in previous in-person studies, the latter of which were simply sitting in the same room as their partner. As such, participants in the current study might have wondered if there were any distractors in their partner’s environment they were unaware off, for example behind their partners’ screen. Not having access to this peripheral information about a partner’s environment might make it more difficult to infer whether joint attention with a partner has truly been established. In the current study, this might have contributed to the joint versus disjoint manipulation being less effective.

Crucially, however, these differences in situational context between our current study and previous in-person studies, that is, (1) not being able to see what happens on a partner’s screen, (2) having more difficulty discerning whether a partner is looking at you or at the stimulus you are jointly attending to, (3) not being able to make eye contact, and (4) not having access to information about a partner’s peripheral environment, are not specific to the current study procedure, but are inherent to all
typical current day forms of video mediated digital communication. Taken together with the high degree of similarity between the current study and previous in-person study with regard to the other aspects of the study, as well as the high ecological validity of the current study in the context of online triadic social engagements (e.g., online lectures), the most plausible interpretation of the current data is that creating social closeness through joint attention in online video mediated interactions is not an effective way of creating social closeness, or at least considerably less effective than in real world in-person interactions.

This does not mean, of course, that the results of the current study imply that creating social closeness through online video-mediated communication in and of itself is impossible. Such interactions might still for example allow individuals to create social closeness through interactions that are not centered around a stimulus on a shared screen, but instead around conversations, especially if in these conversations individuals exchange personal information (Aron et al., 1997) or gossip (Dunbar, 2004). Furthermore, even triadic interactions centered around shared screens might facilitate social bonding if accompanied by more explicit communication during that interaction. Also, online social interactions between individuals characterized by cooperation and interdependence, have been shown to facilitate social bonding (Dabbish, 2008; Depping & Mandryk, 2017; Zheng et al., 2002).

27 Aside from perhaps certain forms of communication through virtual reality.
The current study, however, suggests that connecting through joint attention in itself, one of the fundamental building blocks of creating social closeness in triadic interactions that has been shown to facilitate social bonding in human adults, children, and even, to some extent, great apes (Haj-Mohamadi et al., 2018; Rennung & Göritz, 2015; Wolf, Launay, et al., 2015; Wolf & Tomasello, 2019, 2020a, 2020b), might not work as well, or even at all, in video mediated triadic interactions. This could mean that such online interactions, even when they do contain more explicit forms of communication, might be less effective at forming and maintaining social relationships and social networks than their in-person counterparts. More empirical work on this hypothesis is necessary however, before strong conclusions can be drawn.

Based on the current results it is difficult to draw any strong conclusions of about the role of group size in social bonding through joint attention. There was no difference between the joint and disjoint attention condition in social bonding scores for participants interacting in dyads or groups. Given that the results of the current study with regard to the joint versus disjoint manipulation in dyads are in such stark contrast with the results of previous in-person studies, it is difficult to say whether the current absence of the social bonding effect of joint attention in groups provides meaningful information about the degree to which joint attention is an effective social bonding mechanism in group contexts across different modes of social interaction. Future research should therefore be aimed at directly comparing social bonding through joint
attention in dyads versus groups during in-person interactions to see if social bonding through joint attention in groups generally occurs at similar levels as in dyads.

Nevertheless, one thing is certain. In our current world where digital communication has become an increasingly prevalent part of our everyday social world, we are more and more relying on novel forms of communication to be able to maintain our social relationships, and form new ones. It is therefore crucial to determine exactly how comparable social interactions mediated by digital technology and face to face interactions are in terms of their capacity to create social closeness between people. The stark contrast between the data of our current online study and previous research on social bonding during in-person shared experiences suggests that creating social closeness through video mediated triadic interactions is not as effective as their in-person counterparts. If this is indeed the case, this will have a profound effect on the social relationships and social networks, as well as the overall social cohesiveness in communities and societies where these types of technologies are currently used on a regular basis.
Chapter 6. Social preferences and attitude formation in children’s shared social activities.

Introduction

Humans have a remarkable variety of social activities through which they form and maintain their social relationships in order to satisfy their need to belong (Baumeister & Leary, 1995). But do we simply like to engage in these activities because we enjoy the content of these activities in itself, and is the social bonding effect merely a byproduct? Or is there something about sharing the experience of engaging in an activity with others that changes how much we enjoy that activity? And if sharing an activity makes us feel more positive about what we are doing, does this change our general attitude towards the content of that activity?

To create social closeness with others, humans employ a variety of social bonding activities. Research has shown that humans socially bond by engaging in conversation with partners about themselves (Aron et al., 1997) or others in their group (Bosson et al., 2006; Dunbar, 2004). In addition, humans connect with others through a variety of non-conversational activities, such as playing team sports (Artinger et al., 2006) or cooperative video games (Depping & Mandryk, 2017; Zheng et al., 2002), as well as singing together (Pearce et al., 2015; Weinstein et al., 2016), dancing together in synchrony (Cirelli et al., 2017; Hove & Risen, 2009; Tarr et al., 2015, 2016) and participating in rituals together (Charles et al., 2021; Singh et al., 2020). Yet the question remains why we enjoy engaging in such social activities that much. Are these activities
simply enjoyable? Do we just like to dance and to sing in general, incidentally creating social closeness when we do these things with others? Or do we specifically enjoy them because of their social nature?

Humans seem remarkably motivated to act together with others. Research with human children and great apes showed that, when choosing between collaboratively or individually pulling two ropes to gain access to a reward, children preferred to act together whereas great apes preferred to pull both ropes by themselves (Bullinger et al., 2011; Rekers et al., 2011). This preference for cooperation versus acting alone might explain why we enjoy certain types of social bonding activities, specifically those that rely heavily on behavior coordination, communication and cooperation, such as playing team sports or games (Artinger et al., 2006; Dabbish, 2008; Depping & Mandryk, 2017; Zheng et al., 2002).

However, some of our social bonding activities require only very minimal forms of coordination, communication and cooperation, such as watching a movie in a theatre together. In fact, research with human adults, children and great apes, has shown that even something as minimal as jointly attending to something creates social closeness between individuals (Haj-Mohamadi et al., 2018; Rennung & Göritz, 2015; Wolf, Launay, et al., 2015; Wolf & Tomasello, 2019, 2020a, 2020b). It has been suggested that the reason even such minimal social interactions create social closeness is that during such interactions participants engage in shared cognition, making inferences about the degree
to which their mental states are shared with that of their interaction partners. If so, then, one potential explanation for why humans are so motivated to engage in these kinds of social activities together is that they enjoy the experience of engaging in shared cognition during the interactions these social activities provide.

If engaging in shared cognition is rewarding in itself, we would be more motivated to seek out and stay engaged in activities that create opportunities to engage in shared cognition. But does engaging in a shared activity also change how we feel about the content of that activity beyond a social context? Recent studies with adults suggest that perceptions and attitudes towards the content of an activity change when engaging in this activity together. For example, memories of shared experiences have been shown to be easier to recall than similar experiences that were experienced alone (Eskenazi et al., 2013). Furthermore, several studies have shown that experiencing stimuli by looking at them or tasting them together as opposed to alone causes attitudes and emotions towards these stimuli to become more extreme (Boothby et al., 2014, 2017; Shteynberg, Hirsh, Apfelbaum, et al., 2014; Shteynberg, Hirsh, Galinsky, et al., 2014).

Although there is some indication that joint attention shapes adults’ attitudes towards stimuli, this question has remained largely unexplored for children. There is, however, some evidence to suggest that joint attention does affect the way children process stimuli. Two studies showed that watching a video together with a parent increased skin conductance levels (i.e., increased arousal) and decreased heartrate (i.e.,
increased information processing readiness) in children between 6-13 (Keene et al., 2019; Rasmussen et al., 2017). As such, it seems that children’s psychological experience of co-attending to something is different than watching that same thing by themselves. However, the question remains whether children’s psychophysiological response to jointly experiencing a social activity together translates into a social preference for engaging in shared versus individual social activities, and/or shapes their attitude towards the content of such activities enough to change their behavior during subsequent individual activities with similar content.

To find out, we conducted a two-part study with 4.5-year-old children. In the first session, we examined whether jointly attending to a moving toy (compared to individually watching the toy) influenced children’s motivation to play with that toy after their partner had left (i.e., content attitude). In the second part, we looked at whether children were more willing to stay engaged in the activity of watching a video when they were jointly attending to this video as opposed to them watching that video by themselves (i.e., social preference).
Methods
Participants and Design\textsuperscript{28}

The experiment had a between subject design for condition (joint attention vs. disjoint attention) and a within-subjects design for the two sessions (content attitude session vs. social preference session). In total 56 participants (4.25 – 4.75-year-old, 32 female) participated in the study. One participant was excluded for refusing to participate and another participant was excluded due to being diagnosed with Autism Spectrum Disorder. In addition, 13 participants were excluded from the content attitude session due to technical difficulties with the toy ($N = 8$), or because children and/or parents did not follow instructions ($N = 5$). Furthermore, one participant was excluded from the social preference session as they refused to engage in this session after completing the content attitude session\textsuperscript{29}. As such, the final sample had 54 participants (21 females, $M_{age} = 4.47$, $SD_{age} = .16$), of which 41 completed the content attitude session (31 females, $M_{age} = 4.47$, $SD_{age} = .16$), and 53 of which completed the social preference session (30 females, $M_{age} = 4.48$, $SD_{age} = .16$)

\textsuperscript{28} The data collection for this study was halted due to the institutional shut down of the university’s Child labs due to the Covid 19 pandemic.

\textsuperscript{29} In addition, for one participant E1 mistakenly brought out the second distractor toy 30 seconds too early. However, as the child did not respond to this toy being taken out, and stayed watching the video, this did not seem to have an impact on the dependent measure, and the data was therefore included in the final dataset.
Procedure

Content attitude session

Upon arrival participants and their parents were brought to a waiting room where Experimenter 1 (E1) gave them a quick overview of what to expect and went over the informed consent form. Next, E1 brought the child and parent into the testing room for the first session (content attitude). Here, Experimenter 2 (E2) was already sitting in a chair reading a book. E1 instructed the child to sit in a chair and the parent to sit in the back of the room facing away from the child and the experimenter. Before going into the room, E1 told the participant and their parent in a whispering voice that they would show the child a toy, but that it was important that everyone in the room was quiet, indirectly explaining why E2 would not be talking to the child. After both of them had sat down, the experimenter took a moving toy from behind the curtain and put it on an elevated surface in front of the child and E2. E1 then turned the toy on so it started moving and went back behind the curtain as to not disturb the manipulation.

Crucially, while the toy was moving, we manipulated joint attention (including common ground) in line with our previous studies on joint attention, common ground and social bonding in children and great apes (Wolf & Tomasello, 2020b). In the joint attention condition, E2 did not respond when the child and parent came into the room. Instead, when the toy was taken out by E1, E2 put away their book, made eye contact with the child in response to the toy starting to move and kept attending to the toy until the toy stopped moving. In contrast, in the disjoint attention condition, E2 made eye
contact with the child when the parent came into the room but did not stop reading when the toy was brought out and did not make eye contact with the child in response to the toy moving. Instead, in this condition E2 kept reading the book for the entire duration of the manipulation. In addition, in this latter condition there was a cardboard divider preventing E2 to see the toy while it was moving (see Figure 12 for an overview of the setup in both conditions).

Figure 12: Set up of manipulation during session 1 (content attitude)

After one minute, the toy stopped moving, at which point E1 came out and, simultaneously, E2 pretended to get a phone call and left the room. E1 then briefly explained which buttons to push to make the toy move, then told the participant that they could play with the toy if they wanted to. E1 then gave the toy and went back behind the curtain. Next, the child was left to play with the toy for 30 seconds without

---

30 The toy in Figure 12 (the windmill) was used for initial piloting. However, due to technical issues with this toy, it was replaced with a toy microwave for the experimental sessions.
interference, after which E1 started bringing out increasingly interesting distractor toys every 90 seconds, being: (1) magnetic building balls and sticks, (2) a big plastic truck with a small car in it, and (3) a marble run. Each time a new toy was brought out, E1 said to the child: “you can play with this toy here, or with that toy over there”, to make it clear to the child they could choose to play with whatever toy they wanted, but could not bring toys from one area to another. In total the child played in the room for 5 minutes (i.e., 30 seconds + 3 times 90 seconds), after which E1 took the child and the parent back to the waiting room. Here, the child was allowed to play for 15 minutes to provide a break between test parts (to reduce carry over effects) and to give E1 and E2 time to get the room ready for the next session.

**Social preference session**\(^{31}\)

After the 15-minute break E1 took the parent and child back into the experimental room for the social preference session. In this session the child was asked to sit facing a computer screen on which E1 would start a video once they had sat down. Aside from this difference, most of the procedure was similar to the procedure of the content attitude procedure: E2\(^ {32} \) was already sitting when the parent and child came in and the parent was asked to sit in the same spot as during the content attitude session.

---

\(^{31}\) This part of the study procedure was an honors thesis project by Julia Thielhelm

\(^{32}\) Children engaged in both parts of the two-part study (i.e., watching the moving toy and watching the video) with the same experimenter. As the amount of shared engagement between the child and the experimenter was kept constant during the content attitude sessions, but not the social preference sessions, participants always went through the content attitude session first.
Next, the child was asked to sit in their chair, after which E1 started the video and remained behind the curtain. The joint versus disjoint manipulation in this session was also identical to that of the content attitude session: in the joint attention condition E2 did not make eye contact when the parent and child came in, but instead looked up from their book and made eye contact with the child in response to the video starting, after which they watched the video until the end. In contrast, in the disjoint attention, E2 made eye contact when the parent and the child entered the room, then went back to attending to their book, and did not look up for the duration of the video. Furthermore, similar to in the content attitude session, E2’s vision of the video was blocked by a divider. See Figure 13 for an overview of the manipulation set up.

![Figure 13: Set up of manipulation during session 2 (social preference)](image)

The crucial difference with the content attitude procedure Jowas, however, that the video the children were watching was 5 minutes long, and that E1 now brought distractor toys out during the video (i.e., during the shared or individual activity).
Similar to in the content attitude procedure, each time a new toy was brought out, E1 said to the child “you can play with this toy here, or you can watch the video over there”. In this case, E1 brought out magnet tiles after 30 seconds, followed by a truck carrying several blocks, and a marble run at 90 second intervals. Once the full 5-minute video had been played, E1 and E2 took the child and parent back to the waiting room, provided a debriefing and gas compensation to the parent, and gave a prize (i.e., toy) to the child for them to take home with them.

**Measures and coding**

Two cameras made video recordings of the sessions which were subsequently coded for exclusion criteria (e.g., experimenter error, parent error) by a research assistant. Exclusions were always verified by a second research assistant, who, in the case of a disagreement asked a third research assistant to break the tie. Once a session had been included, a research assistant coded (1) how long it took the child to leave the toy (content attitude) or video (social preference) for the first time, and (2) the total time spent playing with the toy (content attitude) or watching the video (social preference). In both cases, leaving was operationalized as the moment the child passed the corner of the table separating them from the experimenter, and, in case of the content attitude session, had stopped playing with the toy. Twenty percent of all observations ($N = 40$) were coded by a second coder to assess coder reliability, which was high (Intraclass Correlation Coefficient = .997)
Results

The skewness and kurtosis of the residuals for taking leave in both sessions, as well as the total time spent playing with the toy (content attitude session) and watching the video (social preference video) were all between -1.96 and 1.96, meaning that using Bayesian models with default Gaussian priors and likelihoods was warranted (see Figure 14 and 15 for an overview of the data distributions per condition). Furthermore, Bayesian correlations analysis showed no relationship between the taking leave measures of both sessions ($r_{\text{mode}} = -0.024^{33}$, 95% HDI = -0.318, 0.273) or the total time engaged with the toy in the content attitude session and the total time engaged in the video in the social preference session ($r_{\text{mode}} = -0.109$, 95% HDI = -0.385, 0.195). We therefore analyzed the two sessions as independent samples.

---

33 The mode indicates the most likely value of the correlation coefficient given the data. The 95% HDI describes the range of the 95% most likely values of the correlation coefficient. When the 95% HDI overlaps with zero, the parameter of interest (in this case the correlation coefficient) is not deemed credibly different from zero.
Figure 14: Data distributions for Taking leave from the moving toy and Total time spent playing with the moving toy

Figure 15: Data distributions for taking leave from video and Total time spent watching the video
Content attitude

Bayesian T-tests showed no credible difference in the time it took children to first leave the moving toy between the joint attention condition ($M = 98.48$ seconds, $SD = 77.85$ seconds) and the disjoint attention condition ($M = 123.33$ seconds, $SD = 91.57$ seconds; 95% HDI = -30.25, 79.96). Furthermore, we also found no credible difference in the total time children spent playing with the moving toy between the joint attention condition ($M = 134.61$ seconds, $SD = 86.53$ seconds) and the disjoint attention condition ($M = 140.37$ seconds, $SD = 93.57$ seconds; 95% HDI = -52.90, 64.42). See Figure 16 for a graphical representation of the 95% Bayesian HDI’s.

Although some data still has to be collected (i.e., around 30%), it seems unlikely that this additional data will drastically change the pattern of results. Not only are the mean differences for both measures relatively small in the light of the standard deviation for both measures, the trend for both measures is opposite of what we would expect if joint influences children’s attitude towards an object beyond the initial interaction. Both the average time at which children left the moving toy and the total time spent playing with the moving toy are higher in the disjoint attention condition than in the joint attention condition.
Figure 16: Bayesian High Density Intervals containing the 95% most likely values of the condition difference scores for first leave and total time played.
Social preference

We found no credible difference in the time it took children to first leave the video between the joint attention condition ($M = 140.64$ seconds, $SD = 114.94$ seconds) and the disjoint attention condition ($M = 89.94$ seconds, $SD = 93.34$ seconds; 95% HDI = -109.83, 8.43). Similarly, we found no difference in the total time children spent watching the video between the joint attention condition ($M = 155.27$ seconds, $SD = 110.51$ seconds) and the disjoint attention condition ($M = 105.50$ seconds, $SD = 98.51$ seconds; 95% HDI = -108.27, 9.28; See Figure 17).

Despite these difference not being statistically credible, the data shows a trend towards children in the joint attention condition waiting longer to leave the video for the first time and spending more time overall watching the video. The HDI’s only barely overlap with zero and the mean difference is considerable (50 seconds for both measures, out of a possible 300 seconds). The substantial width of the HDI’s (between 117 and 118 seconds) for both variables suggests, however, that there was much variability within the groups. Given that for 11 participants out of the preregistered 64 participants data still needs to be collected, it is hard to say whether this would ultimately lead to a credible difference between conditions. Yet, in the light of the high variability within this measure, the size of the effect will be modest regardless of whether in the final dataset the difference between conditions ends up credible or not.

---

34 NHST equivalent of the analysis: $p = .083$ for first leave and $.089$ for total time spent watching.
Figure 17: Bayesian High Density Intervals containing the 95% most likely values of the condition difference scores for first leave and total time played.
Discussion

It is important to note that the results of the current study are still preliminary, with around 20-30% of the data still to be collected. Nevertheless, these preliminary results suggest that jointly attending to a toy does not change children’s motivation to play with that same object after the joint attention experience has ended and their partner has left the room. The preliminary results with regard to whether children are more motivated to remain engaged in an experience when it is shared with a partner through joint attention were more ambivalent, showing a trend towards children staying engaged in the experience longer in the joint attention than in the disjoint attention condition, but also indicating high levels of within-group variability on the dependent measures.

Despite their preliminary nature, the results do seem, at least in part, to be at odds with previous literature on how shared experiences shape attitude formation in adults. Whereas those studies found that engaging in a shared experience caused participants’ attitude towards objects related to this experience to change (Boothby et al., 2014, 2017; Shteynberg, Hirsh, Apfelbaum, et al., 2014; Shteynberg, Hirsh, Galinsky, et al., 2014), our data showed no sign of participants altering their behavior towards a moving toy depending on whether they had jointly or individually attended to this toy earlier. Thus, our preliminary data suggest that children’s attitude towards the content of an experience is not influenced by this experience being shared or individual, at least
not to the extent that it changes their behavior during subsequent engagement in an activity with similar content.

It seems unlikely that the discrepancy between our preliminary results and the adult literature reflects a fundamental difference between adults and children. Research with children has shown that parental co-viewing does increase arousal and processing readiness, measured through skin conductance and heart rate (Keene et al., 2019; Rasmussen et al., 2017), suggesting that children’s processing of stimuli is indeed altered by whether these stimuli are jointly or individually attended to. Furthermore, research has shown that jointly attending to something can also change attitudes towards an interaction partner in both children (Wolf & Tomasello, 2020a, 2020b) and adults (Wolf, Launay, et al., 2015) in similar ways, suggesting that, in general, children’s attitudes are sensitive to whether or not an experience is shared.

It also seems unlikely that the toys used in the current sample were overall either too exciting, or not exciting enough to elicit play behavior. The data did not show any floor or ceiling effects, but, in fact had high levels of variability across children in both conditions. This high variability, however, may hint towards a potential methodological explanation for our preliminary null-results. It could be that the children’s individual preferences for different kinds of toys and activities had a bigger impact on the dependent measure than the individual preferences of adults for the stimuli used in the adult studies. After all, the children in the current study came into a novel environment
with novel people and were exposed to a wide variety of toys, some of which they might have already been familiar with whereas others might have been novel. In contrast, most, if not all participants in for example the adult study in which participants were asked to eat chocolate were probably already familiar with chocolate (Boothby et al., 2014). As such, the procedure and materials in the current study might have produced more statistical noise than the procedures and materials used in the adult studies.

Another methodological explanation for the divergent results between the current study and previous adult studies might be found in the way attitudes were measured in these studies. Participants in the adult studies were asked to self-report their attitudes towards the content of a shared activity, whereas in the current study attitudes were measured in terms of the motivation to stay engaged during a subsequent individual activity with similar content. It could be that sharing an experience might cause participants to better encode that experience in their memory (Eskenazi et al., 2013), causing them to more vividly remember the emotional impact this experience had on them when explicitly asked about their attitude towards the content of that experience. This increased memory encoding, however, might not be sufficient to change individuals’ behavior during subsequent engagement in similar activities.

Nevertheless, even if sharing an experience does not influence one’s attitude to the content of a social activity in general, we might still enjoy an activity more if we can share that experience with others. If the current trend suggesting that children are more
motivated to stay engaged in a shared activity than in an individual activity with the same content persists, these results might help us understand why humans are so motivated to engage in such a wide variety of social activities with others, but much less so to engage in activities with similar content by themselves.

That said, there might be an alternative explanation for why children would stay engaged longer in a video when they are watching it together. Jointly attending to something in a social setting might create a sense of social obligation or joint commitment to stay engaged in that activity. As such, children might have stayed to watch the video, not because they wanted to, but because they felt they were expected to. It would, however, be surprising if these feelings of obligation and commitment were solely responsible for the trend in our data. After all, research on joint commitment in children has shown that to create a shared commitment between children that is strong enough to alter their behavior during interactions, these commitments often have to be expressed verbally (Gräfenhain et al., 2009; Hamann et al., 2012). As such, it seems unlikely that a single instance of eye contact followed by attending to the same video would have a similar effect. Nevertheless, it is hard to fully exclude the possibility that participants’ feelings of social obligation and shared commitment to some degree contributed to the trend of children watching a video for longer when they were watching it together than when they were watching it by themselves.
Another issue which might explain why we found this trend in the social preference session but not in the content attitude session is the fact that the social preference session was always preceded by the content attitude session, during which participants engaged in the same condition as in the social preference session, with the same experimenter. This means that at the start of the social preference session the participant had already gone through a similar manipulation with that same experimenter for one minute. As such, in the social preference session children might for example have been more comfortable around the experimenter in the joint attention condition than in the disjoint attention condition because they had previously engaged in a joint attention session with that same experimenter during the content attitude session. However, given that (1) the time of the manipulation in the content attitude session was relatively short (i.e., 1 minute), (2) there was a considerable break between sessions, (3) the content attitude and social preference sessions were always run in the same condition, and (4) there was no correlation between children’s behavior between the two sessions, it seems unlikely that any carry over effects from the content attitude session would have qualitatively changed children’s behavior during the social preference session.

Overall, the current preliminary results suggest that the relationships between joint attention, social preferences and attitude formation are more intricate than earlier work suggested. Although children might be more motivated to stay engaged in
activities when engaging in them together than when engaging in them alone, this preference does not necessarily manifest itself in a change in attitude towards the content of that activity that is substantial or stable enough to change children’s behavior during subsequent individual engagement in an activity with similar content. In other words, if the trend in our preliminary data persists, the current work shows that the social nature of shared social activities primarily shapes our motivation to engage in these activities together, but not our motivation to engage in such activities per se.
Chapter 7. The development of the liking gap: children over 5 think that partners evaluate them less positively than they evaluative their partners\textsuperscript{35}

Introduction

Although people generally tend to hold unrealistically favorable views about themselves and their individual abilities (e.g. Alicke, 1985; Kruger & Dunning, 1999), this optimism does not extend to our perception of our social lives. We tend to think that others have richer and more active social lives than we do (Deri et al., 2017), are more connected to others than we are (Whillans et al., 2017), and have less desire to connect with us than vice versa (Epley & Schroeder, 2014). This social skepticism not only affects our holistic perception of our social world, but also influences our perception of individual social interactions in our daily lives.

Boothby et al. (2018) provided an especially striking demonstration of this skepticism. After two strangers interacted with one another briefly, both of them reported liking the other more than they thought the other liked them, a social illusion referred to as the Liking Gap. This tendency to underestimate how positively one is evaluated by another individual even persists beyond the initial interaction, and can be found in established social relationships a year after first meeting the other (Boothby et al., 2018). Boothby et al. (2018) proposed that the Liking Gap is the consequence of

\textsuperscript{35} This chapter has been accepted for publication in Psychological Science (Wolf et al., Accepted)
individuals during interactions constantly monitoring others evaluating them. In doing so, people worry about being evaluated negatively, perhaps particularly when interacting with a stranger. This anxiety might make negative thoughts more salient and/or relevant, causing individuals to become more skeptical about the impression they make on others.

If the Liking Gap is indeed driven by people’s concern for how they are evaluated by others, then, in childhood, the Liking Gap should emerge in parallel with the emergence of children’s concern for their reputation and the impression they make on others. Children start to, initially implicitly, manage their reputations around 5 years of age (Engelmann & Rapp, 2018; Silver & Shaw, 2018). When observed by a same aged peer, 5 year olds are less likely to steal stickers (Engelmann et al., 2012), and steal less when they are told that they are being watched by an invisible adult (Piazza et al., 2011). Additionally, 5 year olds refrain from cheating more often if they have a positive reputation to maintain, whereas 4-year-olds do not (Fu, Heyman, Qian, et al., 2016). In addition to shaping their reputation by avoiding behavior they think will be negatively perceived, 5-year-old children also engage in positive behaviors to shape their reputation. When asking five year olds how many stickers to give to another individual (with no cost to themselves), they are more likely to choose the generous option when the recipient could see them than when the recipient could not see them (Buhrmester et al., 1992; Leimgruber et al., 2012). Moreover, similar increases in 5-year-olds’ generosity
has been observed in response to being observed by a neutral, unaffected peer (Fujii et al., 2015). Interestingly, 5-year olds become particularly prosocial when the recipient either can reciprocate later or is part of her (minimal) in group (Engelmann et al., 2012).

There is one study where, under very specific circumstances, 4 year olds have been found to engage in some form of reputation management as well (Rapp et al., 2019). However, in this study the children were in the room with 3 peers, each of whom was asked to donate stickers to another group of children at the same time. The children were separated by dividers and could not see each other, but at the end of the donation period the name of the highest or lowest donor would be displayed on a screen visible to them all. Both 4 and 5-year-old children donated more when they believed the name of the child donating the most stickers would be displayed (as compared to a control condition). However, the experimental set up in this study allowed children to directly compare themselves with their peers, perhaps introducing a competitive or game-like component to the game. Overall, then, the general consensus in the literature is that behavior motivated by reputational concerns reliably emerges at around 5 years of age in most circumstances.

After reputation management starts to emerge at 5 years of age, these skills become more sophisticated over the following years. During this time children gain a better understanding of self-representational motives (Banerjee, 2002; Banerjee et al., 2012). Between 6 and 11 they become more concerned with appearing fair (Shaw et al.,
2014), and modest (Watling & Banerjee, 2007), and gain a better grasp of how to strategically engage in self-representational behavior, doing so especially reliably around age 8 (Aloise-Young, 1993; Banerjee & Yuill, 1999) although the way in which they do so seems to vary per culture (Fu, Heyman, Cameron, et al., 2016). This is also the point during development when children start to engage in competitive altruism with others watching (Herrmann et al., 2019).

If the Liking Gap is indeed driven by reputational and impression management concerns, then one would expect this phenomenon to have similar ontogenetic trajectory; emerging at 5 years of age, followed by a continuous development through middle childhood. To test this hypothesis, we conducted a study in which we examined the Liking Gap in children between 4 and 11 years of age who were initially unfamiliar with one another. Participants engaged in a brief interaction (i.e. building a tower together) after which they were asked in several ways to indicate how positively they felt about their partner as well as how positively they thought their partner felt about them.

**Methods**

**Participants and design**

In total, 260 children or 130 pairs ages 4-12 participated in this study, which was approved by the University’s Institutional Review Board (IRB). Participants were paired based on age (birth dates within 18 months of each other). Participants were tested at a
local museum in a mid-sized US city as well as at various events throughout this same area. We tested 132 males and 128 females composed of 58.0% White, 15.3% Mixed, 9.6% Hispanic, 6.9% Asian, 3.4% Black, and 1.5% Middle Eastern and 5.0% “Did not report”. Approximately 60% of families reported an annual income of 60,000+ however, roughly another 22% of our sample opted out of disclosing this information.

Of the original sample, 5 dyads (i.e. 10 participants) were excluded from analyses because they did not complete the task (N = 1) or because parents reported their child to be diagnosed with autism or a developmental delay (N = 4). In addition, some participants were excluded that did not seem to understand the questions in the dependent measure (N = 5). During data collection it appeared that it was particularly difficult to recruit 12 year olds, resulting in only three 12-year olds in the sample. Due to this limited sample size for this age group, we decided to not include the 12-year olds (N = 3) in our analyses. Finally, one participant extremely disliked their partner, resulting in a Liking Gap score 2.5 times more extreme than the second most extreme score on that side of the distribution. As this made us concerned that something unusual might have happened during the procedure, we analyzed the data with and without this participant, yielding similar results. As the data including this participant made the distribution of Liking Gap Scores highly kurtotic, we report our analyses without this
participant. As such, the final sample consisted of 241 children. Table 6 shows the distribution of Dyad Gender Composition within each group for the final sample.

**Table 6: Distribution of participants per age group over different types of dyads, and the average age difference per age group.**

<table>
<thead>
<tr>
<th>Age</th>
<th>Dyad Gender Composition</th>
<th>Average age difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Girl - Girl</td>
<td>Boy - Boy</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

<sup>a</sup> Negative numbers indicate that the participant was younger than their partner

---

* We initially pre-registered for 36 children per age group (12 per Dyad Gender Composition), in line with the sample of the original Boothby et al., (2018) paper. Due to limited availability of participants in the same age group that did not know each other, we decided to allow participants to be one age group apart (and accounted for age difference in the analysis). In the later stages of data collection this, however, lead to a situation in which we ended up with dyads in which one participant belonged to a group for which sufficient data had already been collected, whereas the other participant belonged to a group for which data collection was still ongoing. As we had no reason to belief that collecting more data than specified in the preregistration for this reason would meaningfully change the results, we decided to, rather than sending these children away or excluding them post-hoc, include them in the dataset. As such, our new sampling goal was to recruit at least 12 participants per cell (instead of exactly 12 participants). In the final stage of our data collection, our data collection was interrupted by the institutional shut down of all data collection due to the COVID-19 pandemic. We therefore decided to cease data collection indefinitely (before analyzing the data). Open Science Foundation preregistration link: https://osf.io/uvyfn/?view_only=7dcd58b8f0d644bd84669aed4a2bf4af
**Procedure**

While parents, with their child, passed by a testing location where two experimenters were present, they were asked if they were willing to have their child participate in the study. If so, we informed the parent (and the child) that we were looking for two children, around the same age, who did not know each other, to participate in a study in which they were asked to build a tower together for 5 minutes. They were told that after the task an experimenter would ask the child a couple questions about their experience during the task, after which the child would be able to pick out a prize in exchange for their help. Finally, the parents were told that instructions and the interaction (but not the answering of the questions) would be video recorded to make sure no experimenter errors occurred and no outside factors in the testing area interfered with the procedure. If this parent and child agreed to participate, one experimenter started to go through the informed consent and demographic forms together with the parent, while the other would look for a second child in the appropriate age group to complete the dyad.

After consent was obtained from both parents Experimenter 1 (E1) started the cooperative tower building task. The original Boothby et al., (2018) paradigm had participants engage in conversation for 5 minutes. Holding a conversation for 5 minutes without having anything else to do, however, can be challenging for younger children. We therefore presented them with a tower building task specifically designed to serve as
a vehicle for social interaction, while making sure that they did not perceive the task to be competitive, or being judged on performance. To do so, we presented them with the following instructions: “The two of you are going to build a tower together. You can build any kind of tower that you want, but the important part is that you do it together. Okay? You will have 5 minutes to build the tower and at the end of these 5 minutes we will ask you a couple questions about how it went and then get you your prize. Okay you are all set to get started building your tower together.”

At 4 minutes into the play session the children were told that they had one-minute remaining. If participants said they had completed the task before the 5-minute mark, E1 would let them know they still had time left and they should see what they could do to make the tower better. E1 and E2 typically watched some of the play session but also tried to look somewhat preoccupied in order to create an environment for the children in which they felt safe as well as independent with regard to their task. After 5 minutes E1 stopped the camera and directed one child to come with them and the other child to go with E2. The experimenters were both as far from each other and other community members as possible in order to create a private space for the child to answer questions. Next, children answered questions about their experience, after which they were thanked and received their prize.
Measures

We asked each child six questions about their partner in the cooperative task. It was explained to the participants that all questions would be read out loud by the experimenter for all children at all ages (as some older participants indicated that they could read the questions themselves). Answers were given on a seven-point smiley Likert scale (see Figure 18) on a tablet. Furthermore, although the questions were communicated verbally, participants were asked to select the answer on the tablet themselves, even if they also gave a verbal answer.

![Smiley face Likert Scale](image)

Figure 18: Smiley face Likert Scale.

In order to assess the degree to which children comprehended the nature of the Likert scale we used a non-task related practice question first. That is, before engaging the child in the Liking Gap questions, the experimenters first asked the children if they liked ice cream, and to indicate on the Likert scale how much they did. If the child’s verbal answer was not similar to what they had selected on the screen the experimenter went through their answer to make sure the child had indeed selected what the child intended. For example: if a child said they loved ice cream but selected the smallest
smile smiling face, the experimenter would say “So that smiley face says you like it, but you only like it a little, see how the face is just a little bit happy? Is that how you feel about ice cream?”. If the child did not respond to this explanation and kept pressing a button that was incongruent with their verbal answer, the experimenter made a note in the data file so that this participant’s data would be excluded from analysis.

Once the child’s comprehension of the Likert scale was assessed, the child was asked two sets of three questions, based on the initial questions from the Liking gap study for adults (Boothby et al., 2018), but adapted for children. The order in which the sets of questions were presented to the participant was counterbalanced within the dyad. One set of questions was aimed at measuring the participants’ feelings towards their partner after interacting with them. These questions included: “How much do you like the other boy/girl?”, “How much would you like to play with the other boy/girl again?”, and “How much would you like the other boy/girl to be your friend?” (Cronbach’s alpha = .73). The other set of questions assessed how the participants thought their partner felt about them. These questions included: “How much did the other boy/girl like you?”, “How much would the other boy/girl like to play with you again?”, and “How much would the other boy/girl like to be friends with you?” (Cronbach’s alpha = .74). The sets of questions were then compiled into a partner evaluation scale and a scale indicating how the participant thought their partner had evaluated them. Next, a Liking Gap score for each participant was created by taking a
participant’s score indicating how they evaluated their partner, and then subtracting the participant’s score indicating how they thought their partner evaluated them.

**Results**

The standardized residuals of the Liking Gap score controlled for Age (skewness = .55; kurtosis = .775) indicated that the data was appropriate for using standardized Bayesian statistical models with normal likelihoods and uninformed priors in JASP (JASP Team (2020), JASP (Version 0.12.2)) and SPSS 26. As participants engaged in the task in dyads, it was necessary to test whether the data within dyads was consequentially (non)independent. As the Intraclass correlation coefficient (ICC) of the Liking Gap scores within the dyads was lower than .45 (ICC = .104) we concluded that there was no consequential non-independence in the data (Kenny et al., 2006). We therefore proceeded to analyze the data at the individual level.

Given that, across different age groups, there was an unequal distribution of the different types of dyads as well as age differences relative to the participants’ partners (see Table 6 in the methods section), we wanted to make sure that these factors did not confound or moderate the effect of age on the Liking Gap scores. We therefore conducted a Bayesian GLM with Age, Age difference, and Dyad Gender Composition as independent variables and the Liking gap score as the dependent variable. We then compared a basic model with only Age as a predictor variable to models that also
included Age Difference, Dyad Gender Composition and/or their interactions as predictors.

The Bayes Factors in Table 7 (i.e. BF_M and BF_{10}) show that the model with only Age as a predictor fits the data better than models that also include Age Difference, Dyad Gender Composition and/or any of their interactions. This suggests that there was no effect of Age Difference and Dyad gender composition on the liking gap scores, and that these variables did not confound or moderate the effect of Age on Liking Gap scores. Subsequent analyses therefore did not include Age Difference and Dyad Gender Composition as predictor variables.

Table 7: Model Comparison for models including Age, Age Difference and Dyad Gender Composition.

| Models                              | P(M)  | P(M|data) | BF_M  | BF_{10} | error % |
|------------------------------------|-------|----------|-------|---------|---------|
| Age                                | 0.077 | 0.554    | 14.889| 1.000   |         |
| Age + Age Difference               | 0.077 | 0.236    | 3.705 | 0.426   | 0.005   |
| Age + Age Difference + (Age ∗ Age Difference) | 0.077 | 0.101    | 1.354 | 0.183   | 0.004   |
| Null model                         | 0.077 | 0.039    | 0.487 | 0.070   | 0.003   |
| Dyad Gender Composition + Age      | 0.077 | 0.028    | 0.351 | 0.051   | 1.364   |
| Age Difference                     | 0.077 | 0.018    | 0.225 | 0.033   | 0.005   |
| Age + Dyad Gender Composition + Age Difference | 0.077 | 0.012    | 0.148 | 0.022   | 1.214   |
| Age + Dyad Gender Composition + Age Difference + (Age ∗ Age Difference) | 0.077 | 0.005    | 0.062 | 0.009   | 0.930   |
A Bayesian linear regression with Age predicting Liking Gap score showed a positive effect of age on Liking Gap scores (Bayesian estimate of standardized regression coefficient: Mode = .062, HDI\(^{37} = .023 - .101\)), meaning that the tendency of children to report that they liked their partner more than they thought their partner liked them (i.e. the Liking Gap) was more extreme for older children than for younger children. To get a better idea about whether this was a gradual trend, or whether this tendency emerged at a specific age, we compiled participants into different 1-year age groups (e.g. 4 year olds, 5 year olds) and conducted a Bayesian ANOVA with Age Group as a factor and Liking Gap score as a dependent variable. As Table 8 and Figure 19 show, there was no difference between how positively 4 year olds reported to feel about their partner and how positively they thought their partner evaluated them (i.e. the Bayesian HDI of the

\(^{37}\) Bayesian Credible Intervals include the estimated 95% most likely value of, in this case, the standardized regression coefficient. If this interval does not contain zero as a likely value, then the effect is deemed credible.
Liking Gap score contains zero). However, in all other age groups children on average evaluated their partner more positively than they thought their partner evaluated them (although the data from the 11 year olds needs to be interpreted with caution due to the limited sample size in that age group). This suggests that the Liking Gap in children emerges between age four and five.

Table 8: Bayesian parameter estimates for the Liking gap scores at different age groups

<table>
<thead>
<tr>
<th>Age</th>
<th>Mode</th>
<th>Mean</th>
<th>Variance</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>.076</td>
<td>.076</td>
<td>.011</td>
<td>-.126</td>
<td>.279</td>
</tr>
<tr>
<td>5</td>
<td>.324</td>
<td>.324</td>
<td>.011</td>
<td>.118</td>
<td>.529</td>
</tr>
<tr>
<td>6</td>
<td>.281</td>
<td>.281</td>
<td>.008</td>
<td>.103</td>
<td>.460</td>
</tr>
<tr>
<td>7</td>
<td>.384</td>
<td>.384</td>
<td>.011</td>
<td>.175</td>
<td>.592</td>
</tr>
<tr>
<td>8</td>
<td>.316</td>
<td>.316</td>
<td>.010</td>
<td>.124</td>
<td>.508</td>
</tr>
<tr>
<td>9</td>
<td>.481</td>
<td>.481</td>
<td>.014</td>
<td>.251</td>
<td>.712</td>
</tr>
<tr>
<td>10</td>
<td>.583</td>
<td>.583</td>
<td>.019</td>
<td>.315</td>
<td>.851</td>
</tr>
<tr>
<td>11</td>
<td>.500</td>
<td>.500</td>
<td>.047</td>
<td>.076</td>
<td>.924</td>
</tr>
</tbody>
</table>

We also examined if the overall effect of age on the Liking Gap was exclusively driven by its emergence at age 5, or if there was still an effect of age on Liking Gap scores after age 5. A Bayesian linear regression with Age predicting Liking Gap scores on the data for the 5 to 11 year olds still showed a positive effect of age on Liking Gap scores (Bayesian estimate of standardized regression coefficient: Mode = .049, HDI = .002.
- .097). This suggests that even after the Liking Gap emerges at age 5, it gets more extreme as children get older.

Figure 19: Bayesian 95% HDI for Liking Gap scores at different age groups.

Finally, we looked at whether changes in the Liking gap scores over different ages were caused by changes in how the participants evaluated their interaction partner, and/or changes in how they thought their interaction partner evaluated them. A Bayesian linear regression showed no effect of age on the how participants evaluated their interaction partner (Bayesian estimate of standardized regression coefficient: Mode
= -.011, HDI = -.063 , .041), and a credible negative relationship between age and how participants thought their interaction partner evaluated them (Bayesian estimate of standardized regression coefficient: Mode = -.073, HDI = -.127 , -.020). However, when looking specifically at the change between the 4 year olds and the 5 year olds, where the liking gap first appeared, a Bayesian between subjects T-test showed that children’s evaluation of their partner became more positive (Bayesian estimate of between age difference: Mode = .398, HDI = .078 , .717), whereas their perception of how their partner evaluated them did not change (Bayesian estimate of between age difference: Mode = .150, HDI = -.225 , .526). In contrast (and in line with the overall trend), for the 5-11 year olds Bayesian regressions showed no effect of age on the how participants evaluated their interaction partner (Bayesian estimate of standardized regression coefficient: Mode = -.045, HDI = -.111 , .021), and a credible negative effect of age on how participants thought their interaction partner evaluated them (Bayesian estimate of standardized regression coefficient: Mode = -.095, HDI = -.162 , -.028). This suggests that the emergence of the liking gap between 4 and 5 and its widening (i.e., becoming more extreme) between 5 and 11 are driven by different psychological mechanisms.

**Discussion**

The current results show that the Liking Gap emerges at around age 5. At age 4, children did not evaluate their partner more positively than they thought their partner evaluated them, whereas children of 5 years of age or older did. This effect was not
influenced by the gender composition of the dyad, nor the age difference between participants. Additionally, we found that, after its emergence, the liking gap became more pronounced between age 5 and 11.

Additional analyses showed that the initial emergence of the liking gap was driven by 5 year olds evaluating their partner more positively than 4 year olds while their perceptions of how their partner evaluated them did not change between 4 and 5. In contrast the widening of the liking gap (i.e., the gap becoming more extreme) between 5 and 11 was driven by children perceiving their partner’s evaluation of them to become less positive (while their evaluation of their partner remained the same). These findings suggest that there are two important developmental processes underlying the emergence and development of the Liking Gap during childhood.

First, between 4 and 5, evaluations of strangers after a single interaction become more positive. This is the age at which children are coming to understand others’ beliefs (theory of mind), but this social-cognitive development is unlikely to be the key factor in the initial emergence of the Liking Gap as it seems to be driven by a change in children’s attitudes toward others, not their perception of others’ attitudes toward them. A more plausible explanation is that between age 4 and 5 many children are rapidly exposed to interactions with novel individuals, in particular same aged peers, for example, due to kindergarten enrollment becoming mandatory at age 5 in the local area’s public school system. This increased exposure to interactions with peers might contribute to a
decrease in stranger anxiety between these ages, and a consequent increase in the social value of interacting with novel peers and the social relationships that could emerge from them. Yet although 5 year olds seem to evaluate novel individuals more positively than do 4 year olds, their perceptions of how others evaluate them do not match this trend, suggesting that an increase in the social value of interacting with novel individuals does not automatically bring with it a heightened concern for how others are evaluating the self.

Second, after the initial Liking Gap has emerged at 5, children become increasingly concerned with how others are evaluating them (while their evaluation of those others remains the same). This suggests that after its emergence between 4 and 5, the subsequent development of the Liking Gap is primarily driven by increased social concern with others’ evaluations of the self. A likely explanation here is more cognitive, namely, that children’s understanding of their interaction partners’ self-presentational motives increases after 5 years of age (Aloise-Young, 1993; Banerjee, 2002; Banerjee et al., 2012; Herrmann et al., 2019; Watling & Banerjee, 2007), and this shapes their interpretation of their partner’s behavior towards them. Before understanding that others have self-reputational concerns of their own, children might interpret a partners’ positive social behavior during an interaction as a manifestation of their genuine positive attitude towards them. However, as children get older, they start to understand that their interaction partners’ friendly behavior might be motivated by their desire to
make a good impression (e.g., appearing friendly and likeable), and that this behavior is therefore not necessarily indicative of that partner evaluating them positively. As such, children’s increased understanding of others’ self-reputational concerns and impression management strategies might leave them with more uncertainty when trying to infer from their partner’s behavior how positively they are being evaluated.

If this is the case, the question remains how children’s social-cognitive development underlies the emergence of these phenomena related to impression management and reputational concerns during middle childhood (i.e., after age 5). One potentially important social-cognitive skill that develops between age 5 and 11 is the capacity to use second order Theory of Mind, allowing children to process more complex recursively structured social information. For example, at age 6, children have been shown to not only understand and express that others hold beliefs, but also that others hold beliefs about their (i.e., the child’s) beliefs (Grueneisen et al., 2015; Perner & Wimmer, 1985). In a similar fashion, after age 6, children might start to understand that their interaction partner not only holds an attitude towards them (e.g., evaluates them positively or negatively), but that their partner also holds an attitude towards them having an attitude towards their partner. In other words, second order Theory of Mind might be a prerequisite for understanding that others worry about what you think of them. As such, the capacity to engage in higher order Theory of Mind emerging at around 6 might be the social cognitive starting point for understanding others’ self-
representational motives at that age, and thereby thus contribute to the increased social skepticism and widening of the Liking Gap between age 5 and 11 in the current study.

Overall, it seems that the development of the Liking Gap is consistent with previous research on the development of children’s concerns for their reputation and the impression they make on others. Crucially, this suggests that, as Boothby et al., (2018) proposed, the Liking Gap is indeed the result of individuals’ concern for the impression they are making on others during interactions, perhaps especially with strangers. It appears that around the age at which children begin to worry about their reputation and start to manage the impression they make on others, their perception of how others are evaluating them become more negative relative to how positively they are evaluating their partner. As such, children from 5 years on and adults alike might experience the social anxiety inherent to interacting, and thus being evaluated by others, potentially causing negative thoughts to become more salient and relevant, which, in turns, causes individuals to become more skeptical about what others think of them.

The question remains to what extent these findings can be generalized to different populations. Although recruitment at a museum allowed us to include a broader audience than typically the case in laboratory samples, the majority of parents still indicated to be at least middle class in terms of SES. More broadly speaking, the research was conducted in a US mid-sized city, meaning that one should be cautious when generalizing the existence and development of this phenomenon to non-Western
populations (Henrich et al., 2010; Nielsen et al., 2017). In addition, the question remains to what extent individual differences moderate the existence of the liking gap. It is plausible that individual differences that impact the way in which children engage in, or appreciate interactions. In this regard, it would be particularly important to look at the relationship between individual differences in children’s Liking Gap and socially anxious dispositions and attachment styles. It is not implausible that in some cases shyness, social anxiety and/or insecure attachment could be a manifestation and/or consequence of a relatively high discrepancy between how much children like others, and how much they think others generally like them back.

Furthermore, as the task (and probably dependent measure) of the current study were not designed to be suitable for individuals older than 11-12, the question remains whether how the Liking Gap develops through late childhood into early adolescence. Research suggest that children’s and adolescents’ concern for reputation and status within their peer groups peaks at early adolescence (LaFontana & Cillessen, 2010). If future research on the development of the Liking Gap specifically in these age ranges found a similar peak in early adolescence, this would further consolidate the relationship between the Liking Gap and individuals’ concerns for their reputation and the impression they make on others.

Overall, the current results demonstrating the development of the Liking Gap through early and middle childhood illustrates the emergence of humans’ concern for
the impression they make on others during social interactions. It appears such concerns profoundly shape our social reality even from an early age on, by influencing the way we think others evaluate us, and thus, ultimately, shaping the way in which we evaluate ourselves.
Chapter 8: Conclusions

Humans’ seemingly unique ways of social bonding have long been proposed as an explanation for how we are capable of forming and maintaining large and complex social networks. Yet although much research has been done on which social activities contribute to social bonding, the psychology underlying our capacity to connect through these social activities has remained remarkably unexplored. This is surprising, as exploring this part of our psychology is not only key in understanding the fundamentally social nature of our species, but also pivotal in our efforts to maintain and improve our mental and physical wellbeing.

The research in this dissertation provides insight into how humans’ unique shared cognition (1) shapes the way we connect with others, (2) does so in early childhood, (3) is to some degree shared with the great apes, our closest living evolutionary relatives, (4) but seems to work differently for humans than for great apes, and (5) is not effective in creating social closeness through video-mediated online social interaction in dyads or groups. Furthermore, (6) engaging in shared cognition does not change children’s attitudes to the content of shared experiences, but might cause them to have a preference for doing things together as opposed to alone. Finally, (7) shared cognition also has a darker side, which, in middle childhood, manifests itself in the development of a Liking Gap, a social illusion in which children think that, after a brief interaction, their partner evaluates them more negatively than they evaluate their
partner. Together these findings demonstrate the pivotal role of shared cognition in humans’ social world, and its foundation in ontogeny and phylogeny. After a brief summary of the previous empirical chapters, this final Chapter will discuss the theoretical and practical implications of the current work, as well as novel research questions and directions for future research.

**Chapter summaries**

**Chapter 2. Co-attending and social closeness in young children**

This chapter described research showing that 2.5-year-olds were more willing to interact with a novel adult experimenter after this experimenter had watched a video with them than when this experimenter, sitting in the same place, was unable to see the video screen and was reading their own book. This effect was found both when the nonverbal communication between the experimenter and the child was set free to allow for natural joint attention interactions to occur (Study 1), and when experimenters were prohibited from looking to, and nonverbally communicating with the child (Study 2). A similar procedure with 1.5-year-olds did not yield any results. The question remains, however, whether this reflects a developmental trend, or if this is simply a methodological issue. Overall, these results show that young children already connect with others through shared cognition, suggesting that this capacity is fundamental to our species.
Chapter 3. Co-attending and social closeness in great apes

The research described in this chapter shows that great apes, specifically chimpanzees (Pan troglodytes) and bonobos (Pan paniscus) behave more socially towards individuals with whom they watched a video together. In Study 1, chimpanzees and bonobos approached a novel experimenter faster after this experimenter had watched a video with them than when the screen was turned away from the experimenter who was reading their own clipboard instead. In Study 2, pairs of chimpanzees spend more time in the same part of the room after watching a video on the same screen than after watching a video on different screens. Together, these studies suggest that at least some parts of the human capacity to create social closeness through shared cognition is shared with other animals through common descent, and thus has deeper evolutionary roots than previously suspected.

Chapter 4. Joint attention, common ground and social closeness in young children and great apes

This chapter describes a study comparing 2.5-year-old children and great apes (i.e., chimpanzees and bonobos) in how a partner’s attempt to create common ground about an experience being shared influences the subsequent willingness to interact with that partner. In both conditions, for both species, participants watched a video together with an experimenter. In one condition, however, the experimenter attempted to make eye contact in response to the video appearing on the screen in order to create common ground about the fact they were watching the video together. In the control condition
the experimenter made no eye contact in response to the video appearing on the screen, meaning that in this case the participant was merely observing the experimenter watching the same video they were watching. In the latter case, the experimenter made eye contact during a different part of the procedure, unrelated to the video, to control for the general effect of making eye contact on social bonding.

Results showed that children, but not great apes were more willing to approach this novel experimenter when this experimenter had attempted to create common ground about watching the video together than in the control condition. These results suggest that, although humans and great apes share the capacity to create social closeness by watching something together in close proximity, only humans create additional social closeness by establishing common ground about the fact that their experience is shared. This study highlights a nuanced, yet potentially important difference between human and great ape social cognition that might help explain why humans engage in such unique social bonding activities, and are thus able to form and maintain larger social networks than other primates.

Chapter 5. Connecting over the internet: social bonding through joint attention in online video-mediated dyadic and group interactions

In this chapter we examine whether the social bonding effect of joint attention found in previous studies (including those in Chapters 2-4) also operates in the context of online video mediated triadic interactions. Additionally, we also explore whether the effectiveness of joint attention as a social bonding mechanism in these types of
interactions is moderated by group size. Our results show, both in the dyads and in the larger groups, no difference in social bonding scores between the joint attention and disjoint attention condition. As these null-results, even for the dyads, are in stark contrast with previous work on joint attention and social bonding in in-person social interactions, it is difficult to draw any strong conclusions from this data about the effect of group size on connecting through joint attention in general. More research in which connecting through joint attention in dyads and groups is directly compared is necessary before doing so. The current results, however, do suggest that, overall, engaging with others in joint attention during online video mediated interactions is not an effective way of creating social closeness with others, or at least considerably less effective than in in-person face to face interactions. This bears important implications for the formation and maintenance of social relationships and social networks as well as for the social cohesion of our communities and societies as a whole in our current day social world, in which an increasing amount of social interactions take place through digital communication.

Chapter 6. Children’s social preferences for joint attention activities, and the role of joint attention in the formation of attitudes to their content.

In this chapter we describe a study aimed at understanding whether engaging in a shared social activity (1) increases children’s motivation to stay engaged in that activity and (2) changes their attitude to the content of that activity. The preliminary
data in this chapter suggests that that engaging in a shared social activity does not change children’s attitude towards the content of that activity, at least not enough to change their behavior during subsequent exposure to individual activities with similar content. Furthermore, preliminary results also showed no increased motivation in children to stay engaged in a shared activity (i.e., watching a video together) compared to an individual activity (i.e., watching a video alone). However, this preliminary data did show a trend towards children staying engaged longer when jointly attending to a video, which, if this trend persists into the final dataset, might help us explain why humans are so motivated to engage in shared social activities.

**Chapter 7. The development of the Liking Gap**

In this chapter we describe a study on the darker side of the role of shared cognition in social bonding. Specifically, we address the development of the Liking Gap, a social illusion in which, following an interaction, individuals think their interaction partner evaluates them more negatively than they evaluate their interaction partner. Our results show a Liking Gap emerging at around 5 years of age, which is around the same age children become concerned with their reputations and the impression they make on others. Furthermore, we find that the Liking Gap becomes increasingly more intense in older children, which is in line with developmental work showing children’s increased understanding of others’ impression management concerns and reputational strategies at these ages.
Implications and future directions

Joint attention and social bonding in humans

The research in the current dissertation bears important theoretical implications for research on the evolution of human social bonding. The effect of minimally shared interactions on children’s willingness to interact with novel adults (Chapter 2) demonstrates how shared cognition facilitates social bonding even early in ontogeny, suggesting that this way of forming and maintaining social relationships is fundamental to our species’ sociality. Furthermore, the current research shows that some elements of this social bonding mechanism are shared with other species (Chapter 3), but also suggests that there are certain aspects of this mechanism, in particular the role of establishing common ground about an experience being shared, that might be unique to humans (Chapter 4).

As such, the current results might highlight a common denominator in the myriad of different social bonding activities we find in human societies around the world. Social activities such as playing team sports (Artinger et al., 2006) and (video) games (Dabbish, 2008; Depping & Mandryk, 2017), making music together (Pearce et al., 2015; Weinstein et al., 2016), dancing together (Cirelli et al., 2017; Hove & Risen, 2009; Tarr et al., 2015, 2016) and participating in rituals together (Charles et al., 2021; Singh et al., 2020) all require participants to behave in a way that allows fellow participants to make shared cognition inferences relatively easily. First, as most of these activities are
highly coordinated (with some even requiring behavioral synchrony) it is relatively easy for participants to make inferences about the mental states of their fellow participants (e.g., attention, or intentions) and infer that at least some of these mental states converge with- and are similar to their own. In addition, all of these previously mentioned social activities also provide opportunities to establish common ground about these mental states being shared in a variety of different ways, such as by making eye contact while making music, cheering together when a goal is scored in a football match, or making character related comments during a movie. As such, human social activities are particularly suited as vehicles for social bonding. They naturally create social interactions between participants in which these participants are prone to use shared cognition, causing them to feel closer to their fellow participants.

The same applies to social bonding through linguistic communication in conversations, for example by exchanging personal information (Aron et al., 1997) or gossip (Bosson et al., 2006; Dunbar, 2004). In fact, linguistic communication is a particularly effective form of creating shared mental states. Language allows individuals to fairly precisely infer their partner’s mental states, and thus the similarity between their mental states. Furthermore, through language, partners can create a point of mental state convergence on abstract or conceptual content being expressed (rather than being limited to sharing mental states about things that are physically observable in the real world; O’Madagain & Tomasello, 2019). Finally, language is also particularly well
suited to create common ground in conversation (Clark & Brennan, 1991; Tomasello, 2010a, 2014; D. Wilson & Sperber, 2002), as language’s capacity to confer information about abstract concepts allows individuals to directly verify and affirm the degree of mutual awareness of a mental state being shared. In this light, it is not hard to see why engaging in conversations about oneself or others is such an effective way of social bonding (Aron et al., 1997; Bosson et al., 2006; Dunbar, 2004).

Aside from generating new ways to create social closeness with others, shared cognition might also enrich evolutionary older social bonding mechanisms, such as touch. Although many species create social closeness through touch, humans’ shared cognition allows them to mentally converge on the source of their mutual tactile sensation, and also create common ground about this sensation being shared. Similarly, shared cognition might enhance the social closeness resulting from being in close physical proximity with others, simply by having individuals creating common ground with one another about being physically close to one another.

**Contextual flexibility**

Chapters 2-4 showed that the capacity to connect with others through joint attention during shared experiences emerges early in childhood and is shared, at least in part, with our closest evolutionary relatives, the great apes, suggesting that this capacity is deeply ingrained into the psychology of our species. The results of Chapter 5, however, suggest that there are limits to this capacity, showing that creating social
closeness through joint attention in online video mediated triadic interactions is (at least partly) impaired.

To better understand the nature of this impairment it is important to take a closer look at the differences between online video-mediated social interactions and in-person social interactions in the light of the degree to which these contexts afford individuals to make shared cognition inferences. It seems plausible that not being able to see exactly what happens on a partner’s screen or in their peripheral environment outside the camera’s visual field creates more uncertainty with regard to what exactly a partner is attending to. In other words, in video-mediated social interactions it is more difficult to make inferences of convergence, which, in turn, might impair the elicitation of feelings of social closeness.

In a similar fashion, the impossibility of making eye contact during video-mediated social interactions removes a commonly used source of information to make common ground inferences. Although eye contact is not the only way of creating common ground, removing these common automated low effort social signals from the communicative repertoire in video-mediated online social interactions might create more uncertainty about whether there is mutual awareness about an experience being shared, forcing individuals to rely on more effortful explicit communication. This uncertainty about the creation of common ground might, at least in part, negate the additional social closeness that common ground would have elicited in a face to face
setting during an otherwise similar experience. Furthermore, having to resort to more
effortful communication to reduce this uncertainty might also make these social
interactions feel more effortful, and thereby, perhaps, less rewarding, potentially further
reducing the social closeness created during these social interactions.

The results from Chapter 5, taken together with previous research on social
bonding through joint attention in in-person social interactions, highlight how
differences in the situational context of a shared experience might affect its potential as a
social bonding opportunity. It seems plausible that the effectiveness of shared
experiences as a social bonding mechanism are, at least in part, determined by the extent
to which different contexts and environments afford individuals to send and receive the
information necessary to make shared cognition inferences while interacting. Future
research should therefore be aimed at further exploring these contextual limitations of
joint attention as a social bonding mechanism, specifically in contexts in which making
shared cognition inferences is more difficult.

**Group size and social bonding**

Although the results of Chapter 5 did not show a social bonding effect of joint
attention in a group setting, it is difficult to draw any strong conclusions about whether
connecting through joint attention in shared experiences is possible in general. As we
also did not find any effect of joint attention on social bonding in dyads in this online
setting, the lack of effect for participants who were interacting in a group setting might
simply be caused by the mode of interaction (i.e., video chat) rather than the amount of individuals participating in the interaction. Thus, to get a more definitive answer to the question how effective creating social closeness through joint attention is in a group setting, future research should be aimed at directly comparing the social bonding effect of joint attention in dyads and in groups during face to face interactions.

**Shared cognition and long term social bonding**

Additionally, it is important to note that although social bonding is a process that generally spans a multitude of social interactions between individuals, the scope of the current studies has been to assess social closeness immediately following a single interaction. It is therefore important for future research to examine the role shared cognition plays in long term social bonding. Does the social closeness generated in shared experience carry over from that initial experience to future interactions? In this regard, one particularly important hypothesis to explore in future research is that shared experiences not only create momentary social closeness, but also shared memories which, in future interactions, can be used to re-elicit social closeness between individuals, for example by jointly reminiscing about them.

**Joint attention and social bonding in human children**

The results of the current dissertation described in Chapter 2 and Chapter 4 showing that even young children already connect through joint attention might provide a solution for an issue which evolutionary developmental psychologists have
struggled with for several decades. It has been argued that the cognitive capacities underlying shared cognition evolved in humans to facilitate cooperation that became necessary for survival in a rapidly changing ecology (Tomasello et al., 2012). That is, around the time when the human genus first emerged, human ecology experienced a shrinkage in the prevalence of closed forests which changed the human environment into a more grass-land based ecology (Bobe et al., 2002) and caused a radiation of terrestrial monkeys (Elton, 2007). This relatively sudden change in the environment and increased competition with other species made it more difficult to obtain resources that were previously easily acquired individually, forcing humans in an evolutionary niche in which a significant proportion of food acquisition was dependent on sources that could only be acquired through cooperation.

If, however, shared cognition evolved for the purpose of collaboration, then why do these capabilities emerge so early in childhood? Already at 12 months, children start to engage others in joint attention through pointing, and do so, at least in part, because they want to share attention to things they find interesting (Behne et al., 2012; Chen et al., 2009; Cochet & Vauclair, 2010; Liszkowski et al., 2004, 2007b; Murphy, 1978; Tomasello, Carpenter, et al., 2007). This means that there is a significant proportion of joint attention engagements initiated by the child that does not seem to be related to the acquisition of resources. Moreover, it seems unlikely that the emergence of these capabilities this early in ontogeny are related to cooperation or collaboration. After all, at
12 months, children’s limited motor capabilities makes them unlikely collaboration partners. In fact, research has shown that the motivations and capabilities to engage in mutualistic collaborative interactions appear between 2 and 6 years of age (Brownell et al., 2006; McAuliffe et al., 2015; Warneken et al., 2011), which is considerably later than when early manifestations of shared cognition (e.g., joint attention, pointing) emerge.

**Joint attention, social bonding, and the cooperative breeding hypothesis**

Some scholars have therefore argued that although shared cognition might have initially evolved in humans for the purpose of facilitating cooperation, its early emergence in ontogeny must have been driven by a different set of environmental pressures that occurred somewhere in our more recent evolutionary history and was specifically impactful for young children. One account that has been proposed in this regard is the cooperative breeding hypothesis (Burkart et al., 2009b; Hawkes, 2014; Hrdy, 2007b). This idea entails that part of humans’ unique social cognition migrated down in ontogeny due to novel selection pressures in human children resulting from changes in parenting practices in our recent evolutionary history.

Compared to phylogenetically proximate species such as the great apes, the human species is characterized by slow physical and endocranial maturation (Coqueugniot et al., 2004). This means that human children take relatively long to become independent from their parents in terms of predator protection and resource acquisition, and thus strain parental resources for a much longer period of time than in
other species. This prolonged developmental trajectory in our species would result in relatively larger birth intervals and lower reproductive success, and would therefore be disadvantageous from an evolutionary point of view, unless the costs of parental care could somehow be reduced (Kaplan et al., 2000).

To do so, humans started living in cooperative breeding groups, meaning that the costs of childrearing became shared among multiple individuals, raising their offspring in small groups with only a few adults taking care of multiple, often unrelated infants, freeing up time for the other adults to engage in for example foraging or hunting. These cooperative breeding groups posed a novel selection pressure on human children, as they were suddenly required to compete with their peers over the attention of only a few caretakers, requiring them to develop the social cognitive skills necessary to monitor these caretakers, interpret their behavior and elicit solicitude (Burkart et al., 2009b; Hawkes, 2014; Hrdy, 2007b). Children that were able to use these social cognitive skills to create social closeness with a caretaker from a distance would have been able to secure protection and resources more effectively than peers who had not yet developed these skills, and would have therefore been more likely to survive into adulthood.

In this light, the current findings showing that young children create social closeness through joint attention indeed suggest that at least part of the function of shared cognition in young children is enabling them to create social connections with adults. However, the children for whom we found this effect in Chapter 2 and Chapter 4
were already 2.5 years old. Moreover, in Chapter 2, the 1.5 year olds did not show an effect, potentially refuting the claim that young infants use their newly developed joint attention skills to create social connections with adults around them. Yet the null results for the 1.5 year olds in Study 1 of Chapter 2 might have been a methodological artefact. The question is whether for example an approach measure is the best way to measure social closeness in an age group in which there is still considerable variability in locomotive capabilities. Furthermore, although 1.5-year-old children can follow the gaze of others behind barriers, suggesting that they understand what others can see even if they cannot (Moll & Tomasello, 2004), there were certain aspects of the current study that might have made it too difficult for these infants to correctly infer what their partner could see and/or was attending to. After all, children were (1) watching a video in a strange environment (2) with a parent behind them and (3) an experimenter next to them (4) whose vision might or might not have been blocked by a small divider and (5) who responded affirming to their communicative bids even if that experimenter could not see the video (6) but otherwise remained silent and focused on the screen. All of these aspects of the procedure in Study 1 of Chapter 2 might just have made it too difficult for this age group to really comprehend the nature of the interaction they were meant to engage in.

As such, the current results in 2.5 year olds provide a starting point for future research in which the social bonding effect of joint attention in younger children can be
explored using procedures better suited to this age group. For example, to make sure younger children are paying attention to the experimenter one could create an experiment in which an experimenter sits opposite a participant, while the stimulus to which they can jointly attend are presented at their sides, or behind the experimenter, with the experimenter clearly changing their gaze in response to the stimuli appearing. This set up might make it easier for children to infer what exactly the experimenter did and did not see, while drawing the attention to the experimenter by putting them in the center of the child’s visual field.

Additionally, future studies measuring social bonding in younger age groups could explore different behavioral measures of social closeness. One simple adaptation of the social bonding measure used in the current studies that might make this measure more suitable for a younger age group is removing the part where children have to approach over a short distance (as this measure is sensitive to variability in locomotive capacities), but instead simply offer a toy within a child’s reach and measure how long it takes for a child to grab it. Another possibility is starting with the experimenter, the child and the parent close together in an area with toys, followed by the parent at set intervals putting more distance between themselves and the child. One could then measure at which distance the child becomes uncomfortable and/or moves over to the parent. This way, the child is not as dependent on their locomotive abilities to get the toy as they were in the current studies. With such adaptations we might be able to find a
social bonding effect of joint attention in children younger than 2.5, which would provide stronger support for the cooperative breeding hypothesis than the current studies.

Joint attention and creating social closeness with adults versus peers

In the child studies in Chapter 2 and 4 showing that children create social closeness through joint attention, the experimenter with whom the children created social closeness was always an adult. As such, the question remains whether joint attention is also a mechanism children use to connect with peers, and whether the emergence of this peer bonding mechanism coincides with the development of children’s capacity to use shared cognition to connect with adults. From a cooperative breeding perspective, one would expect children to prioritize forming relationships with adults over forming relationships with peers, at least until they become less reliant on their caregivers for their survival.

There is some evidence from research on pointing and cooperation that suggests that younger children are less receptive to engaging in shared cognition with peers than with adults. Studies have shown that one and two year olds produce more joint attention bids for adults than for peers (Franco et al., 2009), and that two year olds are more inclined to act on an adult point than on a point from a peer (Kachel et al., 2018). Finally, a study on peer-cooperation in one to three year olds found that one year olds’ cooperative behavior with peers was less coordinated and more coincidental than peer-
cooperation in older children. In contrast, at 12 to 18 months children already help and coordinate with adult partners in a variety of ways (Liszkowski et al., 2008; Warneken et al., 2006; Warneken & Tomasello, 2006). It is, however, unclear whether this discrepancy is rooted in young children’s lack of motivation to engage with peers, or if their capacity to engage in social interactions is so limited at this age that they can only effectively engage in such interactions when these interactions are scaffolded by adults.

To find out whether the ontogeny of social bonding through joint attention with peers coincides with the development of social bonding through joint attention with adults, future research should be aimed at adapting the current adult-child interaction paradigm to be suitable for interactions between peers. As using an approach based measure of social closeness is difficult in a peer version of this paradigm, as this requires one of the two children to follow very precise behavioral instructions, a social bonding measure suited for a peer context with children can be modeled after the social bonding measure in our study with chimpanzee pairs (Chapter 3, Study 2). For example, pairs of children could either watch a video on the same screen together or on separate screens, after which they engage in a free play session during which they can decide to interact or not. During this play session, one can measure the time spent playing together, as well as verbal communication and physical distance between the participants as proxy measures for social bonding.
Joint attention and social bonding in children: cross cultural variability

The social bonding effect of joint attention has now been established in several studies. However, all of these studies were conducted in Western, Educated, Industrialized, Rich, Democratic (i.e., WEIRD) societies (Henrich et al., 2010), namely in the UK (Wolf, Launay, et al., 2015), Germany (Rennung & Göritz, 2015), and the US (Haj-Mohamadi et al., 2018; Wolf & Tomasello, 2020a, 2020b). Moreover, only two of these studies examine the social bonding effect of joint attention in children (Wolf & Tomasello, 2020a, 2020b), meaning that all our knowledge about the ontogeny of this phenomenon is based on US samples. There is, however, some reason to expect the emergence of social bonding through joint attention to vary across cultures, for example depending on how often children are exposed to joint attention interactions in their early development. It is not implausible that in cultures in which engaging in face to face interactions with children during early childhood is less common than in Western cultures, such as in the BaYaka and Bondongo tribes in Congo (Lew-Levy et al., 2020), social bonding through shared cognition emerges later in ontogeny than in WEIRD countries. To explore the cross cultural variability in the ontogeny of connecting with others through joint attention, future research should aim to replicate the findings of the current child studies in Chapter 2 and 4 in cultures in which practices with regard to early child rearing and early child social interactions substantially differ from that of WEIRD countries.
Joint attention and social bonding in great apes

The results of the current dissertation provide new insight into a long debate about great ape social cognition. Although there is now plenty of evidence showing that great apes can make inferences about the mental states of others (e.g., Hare et al., 2000, 2001; Kano et al., 2019; Krupenye et al., 2016; Schmelz et al., 2011), some have argued that apes are incapable of directly comparing others’ mental states to their own (Tomasello, 2014, 2019). The data from Chapter 3, however, suggests that great apes do compare their mental states with the mental states of others. That is, watching a video with an experimenter (Study 1) and with a conspecific (Study 2) caused great apes to adjust their subsequent social behavior towards this individual, which not only suggests that they are capable of inferring that their attention converges with the attention of a partner, but also that they attribute social value to this convergence. This data thus demonstrates that great ape social cognition is more similar to that of humans than previously thought.

In contrast, Chapter 4 highlights a nuanced, yet potentially important difference between great ape and human social cognition. Human children, but not great apes behaved more socially towards an experimenter when this experimenter attempted to create common ground with them by making eye contact in response to a stimulus appearing on a screen, than when the experimenter did not make such an attempt, and the participant was merely observing the experimenter watching the same video they
were watching. It seems unlikely that the absence of a change in behavior in the apes in response to the manipulation was caused by the apes not being capable of processing the gaze of the experimenter. After all, apes have been shown to understand what others can and cannot see (Bräuer et al., 2007; Hare et al., 2000, 2001), even when an individuals’ line of sight is interrupted by an opaque barrier, as opposed to a transparent window (Okamoto-Barth et al., 2007; Tempelmann et al., 2011). Another issue could have been that the dependent measure might not have been appropriate for a human-great ape comparison, or that the great apes simply did not care about the experimenter in this study because it was a human as opposed to a conspecific. Both of these explanations also seem unlikely as the results of Chapter 3 (Study 1) show that great apes’ response to a slightly different manipulation yielded differences on the exact same dependent measure with a human experimenter.

As such, the most likely explanation for these results is that human children and great apes somehow interpreted the gaze alternation of the experimenter between the screen and the participant differently. This could indicate that perhaps great apes struggle with the recursive nature of the common ground inference that can be made based on this gaze alternation (i.e., the experimenter looking at the participant allows the participant to infer that now both of them are aware that they both know that they are watching the same video). Another possibility is that apes did process the common ground inference correctly, but did not attach any social value to it, at least not to the
point where it changed the way they behaved towards the experimenter. Either way, the current results suggest that, at least in the realm of social bonding, great apes do not seem to be influenced by (attempts to create) common ground the way that human children are, potentially demonstrating a fundamental difference between human social bonding and that of great apes.

In line with a recent opinion piece in Biology Letters (Busia & Griggio, 2020), future research should attempt to further explore the social bonding properties of shared experiences in non-human animals more broadly. First, to get a better idea of the phylogeny of creating social closeness through co-attending and/or establishing common ground, comparative research with multiple species has to be done to track the development of this social bonding mechanism through our evolutionary history. Furthermore, future research should also examine the extent to which different social dynamics within a shared experience facilitates great ape social closeness. For example: does more explicit communication during a shared experience create more social closeness in apes in a way that the eye contact failed to do in Study 2 of Chapter 4? Getting a better understanding of the contextual variability of great apes using (some aspects of) shared cognition to connect with others will not only further our understanding of the degree of similarity between human and great ape social cognition, but might also change our perspective on common great ape social bonding practices.
such as fission fusion traveling and alliance fighting, and could inform the development of novel enrichment activities for zoos and great ape sanctuaries.

**Mechanisms of shared cognition based social bonding**

One important question with regard to the role of shared cognition in social bonding that needs to be addressed in future research is what exactly the mechanism is through which shared cognition creates social closeness. The current work corroborates the idea that even minimal forms of shared cognition such as joint attention creates social closeness between individuals. Furthermore, research has shown that even minimalistic cues of others’ motivation to engage in a shared experience such as gaze following are experienced as socially rewarding (I. Gordon et al., 2013; Schilbach et al., 2010) and positively influences people’s attitudes towards partners that are following their gaze (Bayliss et al., 2013; Frischen et al., 2007). So the question is, why do these kinds of social dynamics play such an important role in our social lives?

With regard to joint attention specifically, one explanation is that it is not the sharing of attention itself that makes us feel closer to others, but that joint attention allows us to infer other mental states being shared. For example, as we primarily attend to things that are relevant to our goals, we might be able to infer a partners’ intentions and motivations from where they are looking. As such, realizing that we are looking at the same thing as a partner might allow us to infer that a partner places the same special value on the thing we are looking at, and thus has goals and motivations similar to ours.
Yet, although this explanation is worthwhile to consider when attempting to identify the social bonding effect of joint attention specifically, it essentially merely displaces the problem, as it does not explain the social bonding effect of shared cognition in general. Even if we connect through joint attention because it allows us to infer other mental states that are shared with our interaction partners, then why does sharing these other mental states allow us to create social closeness? Or, to rephrase the question in behavioral terms: why does engaging in shared cognition with someone make us want to interact with them more in the future?

There are several potential mechanisms underlying the social bonding effect of engaging in shared cognition. First, many have argued that shared cognition in humans evolved to facilitate collaboration and cooperation, which has been crucial to our species’ survival (Tomasello, 2014; Tomasello et al., 2012). In this light, successfully engaging in a shared experience with a partner might allow individuals to infer the prowess of their interaction partner’s social cognitive skills. In other words, a partner with whom one can effectively engage in a shared experience might also be a more successful cooperator in future interactions, making them more attractive interaction partners.

Second, inferring shared mental states with an interaction partner might allow individuals to make inferences about their partner’s willingness to interact with them, cooperate with them and support them in the future, perhaps as part of their social
alliance. When an individual is aware that an interaction partner often follows their gaze and tries to establish common ground about what they are seeing, this signals to this individual that their partner finds them socially relevant and has a positive or even cooperative attitude towards them. As such, social signals associated with shared experiences, such as gaze following, gaze checking, eye contact in response to external stimuli and communicative behavior, can be construed as implicit signals of a partner’s motive to affiliate. This, in turn, would lead an individual to expect more social support and cooperation and less aggression in subsequent interactions, making the prospect of interacting with that partner in the future more appealing.

Third, engaging in a shared experience on a larger scale, with a group, might create a sense of social closeness or belonging to that group, and thereby make individuals more willing to interact with individuals from that group in the future, even if these members did not participate in that initial experience. This specific form of shared cognition operating on a group level has been referred to as collective intentionality (Kern & Moll, 2017; Tomasello & Rakoczy, 2003), and has often been studied in the context of cultural practices, norms and conventions. However, collective intentionality might play a similarly important role in social bonding within (cultural) groups, and the lack thereof between different cultural groups.

Fourth, engaging with others in shared experiences might provide us with information about these individuals that makes it easier to process and interpret their behavior in future interactions. When individuals engage in shared experiences, they
learn about their partner’s mental states, some of which might be relatively stable, such as attitudes, knowledge and long term goals. Learning about a partner’s stable mental traits allows individuals to create heuristics and schemas that they can use to predict and anticipate a partner’s behavior in future interactions more effectively. Importantly, this would cause the processing of that partner’s behavior in future interactions to be less effortful.

This increase in processing fluency during social interactions, caused by learning about a partner during previous shared experiences, might lead individuals to form a more positive attitude towards their partner, as well as the prospect of interacting with them in the future. Processing fluency, that is, the ease with which information is processed, has long been shown to cause individuals’ attitudes towards that information and things associated with this information to become more positive (Alter & Oppenheimer, 2006, 2009; Bornstein, 1989; P. C. Gordon & Holyoak, 1983; Laham et al., 2009, 2009; Reber et al., 2004; Reber & Unkelbach, 2010; Song & Schwarz, 2009; W. Wilson & Miller, 1961; W. Wilson & Nakajo, 1965; Zajonc, 1968). Recently, researchers have also shown that the fluency of processing information during interactions has a similar effect. That is, increased processing fluency during interactions causes individuals to feel more positive about their interaction partner, as well as the prospect of interacting with them in the future (Dragojevic et al., 2017; Dragojevic & Giles, 2016; Koudenburg et al., 2013; Truong et al., 2020).
As such, each shared experience with a partner provides novel information about that partner’s psychology that can be internalized, which causes information processing in each subsequent interaction to be relatively more fluent. This, in turn, makes future interactions with that partner relatively more desirable. Moreover, creating common ground during shared experiences adds to this processing fluency based social attitude formation in a similar way: the mutual awareness of both interaction partners knowing about their partner’s attitudes, preferences and knowledge, makes it easier for both individuals to anticipate their partner’s behavior, interpret their partner’s communication and behave in a way their partner will respond positively towards. As such, establishing common ground with a partner increases processing fluency during subsequent interactions with that partner even further and makes the prospect of these interactions in the future even more appealing.

Finally, this processing fluency based social bonding mechanism can also be applied to social bonding on a group level, through collective intentionality. Shared cognition allows individuals to assume that certain traits (e.g., preferences) or decision rules (e.g., norms) are broadly shared within a social group, making it easier to anticipate and interpret the behavior of people within that group during interactions. As such, interacting with these group members might be (anticipated to be) more enjoyable than interacting with individuals from another group. Hence the argument that
processing fluency plays a major role in facilitating prejudice and intergroup conflict (Lick & Johnson, 2015).

Overall, there are several mechanisms through which shared cognition might facilitate social bonding. Importantly, these mechanisms are not necessarily mutually exclusive, but might co-exist, and even complement each other. To assess the validity of these proposed mechanisms, future research should be aimed at assessing individuals’ perceptions of a partner’s competence and motivation to collaborate in future interactions following a shared experience (compared to a non-shared experience). This will allow us to see whether social bonding through shared experiences is related to individuals’ perceptions of a partner’s perceived competence in social coordination and communication, and a partner’s motivation to cooperate in the future. Furthermore, longitudinal research measuring (perceived) interaction fluency and social closeness at different stages of individuals’ social relationships might help us better understand the nature of interaction fluency as a social bonding mechanism.

**Joint attention, social preferences and attitude formation**

An additional direction for future research should be aimed at better understanding the discrepancy between the adult literature on shared experiences and attitude formation, showing that shared experiences amplify individuals’ attitudes towards the content of that experience (Boothby et al., 2014, 2017; Shteynberg, Hirsh, Apfelbaum, et al., 2014; Shteynberg, Hirsh, Galinsky, et al., 2014) and our preliminary
data from Chapter 6 showing that jointly attending to a toy did not change participants’ behavior towards that toy during subsequent exposure. Future research in adults should measure participants’ attitudes towards the content of a shared versus individual experience by assessing their behavior during subsequent individual exposure to the same or similar content. This would provide insight into whether the attitude change found in previous adult studies actually meaningfully changes adults’ behavior in the long term.

Additionally, future research in both children and adults should aim to more directly assess whether shared experiences are intrinsically enjoyable. The claim from social psychologists that shared experiences are not inherently enjoyable because jointly attending to stimuli with a negative valence causes self-reported attitudes and emotions to become even more negative (Boothby et al., 2014) seems to be flawed. One cannot simply equivocate individuals’ attitude towards the content of a social activity with their enjoyment of this activity occurring in a social setting. One can easily imagine how watching a horror movie together causes this film to be perceived as scarier, but that at the same time, watching this movie together is more enjoyable than watching it alone. To gain a better understanding of whether engaging in shared experiences comes with an inherent psychological reward (relative to engaging in a similar individual experience) future research should for example adopt a procedure in which participants are told that they are about to watch a movie that is likely to elicit negative emotions and
ask them if they would prefer to watch it on their own, or together with someone else, and then ask them afterwards (1) how they felt about the movie, and (2) how they felt about the fact that they watched the movie together/alone. This should give us a more definitive answer as to whether sharing experiences is intrinsically rewarding.

**The Liking Gap**

The current results from our Liking Gap study (Chapter 7) show that even children as young as 5 are sensitive to the Liking Gap illusion. However, it is unclear whether this social illusion occurs specifically in response to engaging in a social interaction with a stranger, or is a more general manifestation of a default way of thinking about how others evaluate us compared to how we evaluate them. Future research should therefore be aimed at assessing whether a Liking Gap can also be found when asking participants how positively they think they will evaluate an interaction partner and how positively they think their partner will evaluate them in anticipation of an interaction.

Another important avenue for future research is exploring the development of the Liking Gap over adolescence. Given that individuals’ concerns for how others think about them peaks during adolescence (LaFontana & Cillessen, 2010), we would expect the Liking Gap to be most extreme during these years and then perhaps decrease in strength over early adulthood. Similarly, future research should be aimed at studying the Liking Gap in socially anxious individuals. One would expect children with social
anxiety to have a more pronounced Liking Gap illusion than typically developing children, as they have been shown to consistently underestimate how much their peers like them (Baartmans et al., 2019). Gaining more insight into the Liking Gap in both of these populations will not only provide insight in how these populations perceive themselves relative to others, but could also inform interventions aimed at helping these populations understand their own Liking Gap, perhaps allowing them to cope more effectively with the social problems they are experiencing as a consequence of their Liking Gap.

Finally, the Liking Gap might also be influenced by individuals’ membership of certain social groups. Research on evaluations of social groups in the US has for example shown that, aside from a strong preference for one’s own group, individuals show a relatively consistent pattern of preferences for religious, age and racial groups that they do not belong to (Axt et al., 2014). Not only will such preferences influence how much individuals from different groups like each other following an interaction, but it will most likely also influence people’s perceptions of how outgroup members evaluate them following an interaction. As such, studying the Liking Gap in individuals following interactions with members of different social groups might help provide important insight into persistent intergroup tensions within societies.
Clinical applications: Shared cognition, social bonding and Autism Spectrum Disorder

One population for which the current research on shared cognition and social bonding might have important practical implications is individuals with Autism Spectrum Disorder (ASD). Children with ASD experience more loneliness and report having lower quality friendships than typically developing children (Bauminger et al., 2003; Bauminger & Kasari, 2000). This suggests that for some reason they struggle with forming social closeness with others more so than typically developing populations.

At the same time, the development of social cognitive skills and social behavior related to shared cognition, such as initiating joint attention, is often impaired or delayed in children with ASD (Charman, 2003; Charman et al., 1997; Dawson et al., 2004; Gomez, 2005; Jones & Carr, 2004; Mundy et al., 1986, 1994). As the work in this dissertation demonstrates the crucial role of shared cognition in human social bonding, impairments or delays in developing these social cognitive skills might have a significant impact on children’s capacity to create social relationships with others. That is, children with ASD might struggle in engaging in shared experiences with others or recognizing when others attempt to engage in a shared experience with them, or they might simply not see the social value of engaging in shared experiences with others altogether. If so, they would miss out on a myriad of social bonding opportunities available to typically developing children, resulting in fewer and lower quality social relationships.
As such, the work in the current dissertation might inform the development of interventions for children with ASD specifically aimed at training them to recognize and capitalize on shared experiences as social bonding opportunities. Additionally, one might supplement conventional conversation training for high functioning children with ASD by specifically training them to refer to past shared experiences they enjoyed as a means to reinforce their social connections with others. Such interventions might make it easier for children with ASD to establish and maintain social relationships with others and, in doing so, increase their mental and physical wellbeing.

**Coda**

Forging and maintaining social relationships is fundamentally important to us as individuals, as well as the human species as a whole. Yet we know remarkably little about the psychological mechanisms underlying some of our most widely used uniquely human social bonding activities. The studies presented in this dissertation not only highlight the importance of shared cognition in social bonding, but also demonstrate that the relationship between shared cognition and social bonding is deeply rooted in the human species, both in ontogeny and phylogeny, and bears important social consequences that shape our social world on a daily basis.
### Appendix A: Supplementary materials Chapter 3

#### Table 9: Demographics and number of pairs and trials subjects participated in
(Chapter 3 Study 2)

<table>
<thead>
<tr>
<th>Name</th>
<th>Year of Birth</th>
<th>Sex</th>
<th>Age</th>
<th>N pairs</th>
<th>N Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asega</td>
<td>1998</td>
<td>Male</td>
<td>20</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Okech</td>
<td>1999</td>
<td>Male</td>
<td>19</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Kisembo</td>
<td>1999</td>
<td>Male</td>
<td>19</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Becky</td>
<td>1991</td>
<td>Female</td>
<td>27</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Nani</td>
<td>2001</td>
<td>Female</td>
<td>17</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Pasa</td>
<td>1999</td>
<td>Female</td>
<td>19</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Bili</td>
<td>1998</td>
<td>Female</td>
<td>20</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Bwambale</td>
<td>2000</td>
<td>Male</td>
<td>18</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Umutama</td>
<td>1996</td>
<td>Male</td>
<td>22</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Cocoa</td>
<td>2007</td>
<td>Female</td>
<td>11</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Indi</td>
<td>1999</td>
<td>Male</td>
<td>19</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Sally</td>
<td>1991</td>
<td>Female</td>
<td>27</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Nkuumwa</td>
<td>1996</td>
<td>Female</td>
<td>22</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Yoyo</td>
<td>1998</td>
<td>Female</td>
<td>20</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Nakuu</td>
<td>2001</td>
<td>Female</td>
<td>17</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Kalema</td>
<td>1996</td>
<td>Male</td>
<td>22</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Baluku</td>
<td>1998</td>
<td>Male</td>
<td>20</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Umugenzi</td>
<td>1997</td>
<td>Male</td>
<td>21</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Medina</td>
<td>2007</td>
<td>Female</td>
<td>11</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Mawa</td>
<td>1996</td>
<td>Male</td>
<td>22</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Namukisa</td>
<td>1998</td>
<td>Female</td>
<td>20</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>
Figure 20: Individual distribution of approach latency, separated by species (Chapter 3, Study 1)
Figure 21: layout of the rooms (Chapter 3, Study 2)
Figure 22: Room division during coding of Same part of the room measure (Chapter 3, Study 2)
## Appendix B: Supplementary materials Chapter 4

### Table 10: Great ape demographics (Chapter 4, Study 2)

<table>
<thead>
<tr>
<th>Name</th>
<th>Sex</th>
<th>Age at time of testing</th>
<th>Rearing history</th>
<th>Species</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bambari</td>
<td>f</td>
<td>19</td>
<td>parent</td>
<td>Chimpanzee</td>
<td>Group A</td>
</tr>
<tr>
<td>Dorien</td>
<td>f</td>
<td>39</td>
<td>hand reared</td>
<td>Chimpanzee</td>
<td>Group A</td>
</tr>
<tr>
<td>Fraukje</td>
<td>f</td>
<td>44</td>
<td>hand reared</td>
<td>Chimpanzee</td>
<td>Group A</td>
</tr>
<tr>
<td>Frodo</td>
<td>m</td>
<td>26</td>
<td>parent</td>
<td>Chimpanzee</td>
<td>Group A</td>
</tr>
<tr>
<td>Lobo</td>
<td>m</td>
<td>15</td>
<td>parent</td>
<td>Chimpanzee</td>
<td>Group A</td>
</tr>
<tr>
<td>Lome</td>
<td>m</td>
<td>18</td>
<td>parent</td>
<td>Chimpanzee</td>
<td>Group A</td>
</tr>
<tr>
<td>Riet</td>
<td>f</td>
<td>42</td>
<td>hand reared</td>
<td>Chimpanzee</td>
<td>Group A</td>
</tr>
<tr>
<td>Sandra</td>
<td>f</td>
<td>26</td>
<td>parent</td>
<td>Chimpanzee</td>
<td>Group A</td>
</tr>
<tr>
<td>Tai</td>
<td>f</td>
<td>17</td>
<td>parent</td>
<td>Chimpanzee</td>
<td>Group A</td>
</tr>
<tr>
<td>Alex</td>
<td>m</td>
<td>19</td>
<td>hand reared</td>
<td>Chimpanzee</td>
<td>Group B</td>
</tr>
<tr>
<td>Daza</td>
<td>f</td>
<td>34</td>
<td>unknown</td>
<td>Chimpanzee</td>
<td>Group B</td>
</tr>
<tr>
<td>Frederike</td>
<td>f</td>
<td>46</td>
<td>unknown</td>
<td>Chimpanzee</td>
<td>Group B</td>
</tr>
<tr>
<td>Hope</td>
<td>f</td>
<td>29</td>
<td>parent</td>
<td>Chimpanzee</td>
<td>Group B</td>
</tr>
<tr>
<td>Zira</td>
<td>f</td>
<td>22</td>
<td>parent</td>
<td>Chimpanzee</td>
<td>Group B</td>
</tr>
<tr>
<td>Gemena</td>
<td>f</td>
<td>14</td>
<td>parent</td>
<td>Bonobo</td>
<td>NA</td>
</tr>
<tr>
<td>Jasongo</td>
<td>m</td>
<td>29</td>
<td>parent</td>
<td>Bonobo</td>
<td>NA</td>
</tr>
<tr>
<td>Kasai</td>
<td>m</td>
<td>7</td>
<td>parent</td>
<td>Bonobo</td>
<td>NA</td>
</tr>
<tr>
<td>Kuno</td>
<td>m</td>
<td>23</td>
<td>hand reared</td>
<td>Bonobo</td>
<td>NA</td>
</tr>
<tr>
<td>Luiza</td>
<td>f</td>
<td>15</td>
<td>parent</td>
<td>Bonobo</td>
<td>NA</td>
</tr>
</tbody>
</table>
Appendix C: Supplementary materials Chapter 5

Table 11: Model comparison parameters for impoliteness models.

| Models                      | P(M) | P(M|data) | BF<sub>M</sub> | BF<sub>10</sub> | error %  |
|-----------------------------|------|----------|----------------|----------------|----------|
| Null model                  | 0.200| 0.543    | 4.748          | 1.000          |          |
| Attention                   | 0.200| 0.217    | 1.109          | 0.400          | 9.234e-5 |
| Group size                  | 0.200| 0.158    | 0.751          | 0.291          | 0.037    |
| Attention + Group size      | 0.200| 0.060    | 0.256          | 0.111          | 1.919    |
| Attention + Group size + Attention * Group size | 0.200| 0.022    | 0.090          | 0.041          | 7.313    |

Table 12: Model comparison parameters for attractiveness models.

| Models                      | P(M) | P(M|data) | BF<sub>M</sub> | BF<sub>10</sub> | error %  |
|-----------------------------|------|----------|----------------|----------------|----------|
| Null model                  | 0.200| 0.630    | 6.820          | 1.000          |          |
| Attention                   | 0.200| 0.182    | 0.887          | 0.400          | 9.234e-5 |
| Group size                  | 0.200| 0.138    | 0.638          | 0.291          | 0.037    |
| Attention + Group size      | 0.200| 0.039    | 0.162          | 0.111          | 1.919    |
| Attention + Group size + Attention * Group size | 0.200| 0.012    | 0.047          | 0.041          | 7.313    |
Table 13: Model comparison parameters for viewing enjoyment models.

| Models                                  | P(M) | P(M|data) | BF_M | BF_10 | error % |
|-----------------------------------------|------|----------|------|-------|---------|
| Null model                              | 0.200| 0.406    | 2.737| 1.000 |         |
| Attention                               | 0.200| 0.357    | 2.225| 0.880 | 5.735e-5|
| Group size                              | 0.200| 0.107    | 0.478| 0.263 | 0.036   |
| Attention + Group size                  | 0.200| 0.094    | 0.417| 0.232 | 0.954   |
| Attention + Group size + Attention * Group size | 0.200| 0.035    | 0.146| 0.087 | 3.175   |

Table 14: Model comparison parameters for video liking models.

| Models                                  | P(M) | P(M|data) | BF_M | BF_10 | error % |
|-----------------------------------------|------|----------|------|-------|---------|
| Null model                              | 0.200| 0.435    | 3.085| 1.000 |         |
| Attention                               | 0.200| 0.360    | 2.250| 0.827 | 5.958e-5|
| Group size                              | 0.200| 0.096    | 0.426| 0.221 | 0.33    |
| Attention + Group size                  | 0.200| 0.077    | 0.332| 0.176 | 1.492   |
| Attention + Group size + Attention * Group size | 0.200| 0.032    | 0.131| 0.073 | 3.200   |
Table 15: Model comparison parameters for watching video again models.

| Models                        | P(M) | P(M|data) | BF_M | BF_10 | error % |
|-------------------------------|------|----------|------|-------|---------|
| Null model                    | 0.200| 0.634    | 6.935| 1.000 |         |
| Attention                     | 0.200| 0.175    | 0.851| 0.277 | 0.037   |
| Group size                    | 0.200| 0.137    | 0.637| 0.217 | 0.033   |
| Attention + Group size        | 0.200| 0.037    | 0.154| 0.058 | 2.390   |
| Attention + Group size + Attention * Group size | 0.200| 0.016    | 0.055| 0.025 | 13.607  |
References


Fu, G., Heyman, G. D., Qian, M., Guo, T., & Lee, K. (2016). Young children with a positive reputation to maintain are less likely to cheat. *Developmental Science, 19*(2), 275–283. https://doi.org/10.1111/desc.12304


229


Mulcahy, N. J., & Call, J. (2009). The performance of bonobos (Pan paniscus), chimpanzees (Pan troglodytes), and orangutans (Pongo pygmaeus) in two


Wolf, W., Nafe, A., & Tomasello, M. (Accepted). The development of the liking gap: Children over 5 think that partners evaluate them less positively than they evaluative their partners. *Psychological Science.*


Biography

Wouter Wolf obtained a BSc in Criminology from Leiden University (The Netherlands), where he subsequently also received his BSc in Psychology in 2011. In 2013 he graduated Cum Laude from the VU University Amsterdam from his MSc in Social Psychology where he published his work on a social media based paradigm to study ostracism in the journal Behavior Research Methods. In 2013 he was also granted a Prins Bernhard Cultuurfonds scholarship for the Sciences to conduct research at the Social and Evolutionary Neuroscience Research Group in the Department of Psychology at the University of Oxford (UK). Here, he published his first paper on joint attention and social bonding in the British Journal of Psychology.

After two years teaching and working in a social neuroscience lab at the VU University Amsterdam, Wouter started his PhD with Professor Michael Tomasello at Duke University in 2016, in part supported by a Fulbright Scholarship from the Netherlands. During his time at Duke, his research was supported by two Lafitte research grants, and for his final year Wouter was awarded the Philip Jackson Baugh Fellowship for research in human development. During his time at Duke, he published two papers on his work on joint attention and social bonding in young children in the Journal of Experimental Child Psychology, as well as a paper on joint attention and social bonding in the Proceedings of the Royal Society B: Biology, which received broad attention from the popular press, leading to interviews with the BBC, the Times, and Science.
Magazine. Finally, he recently also published a paper on the development of the Liking Gap in the journal *Psychological Science*. 