Durham E-Theses

A comparative developmental approach to multimodal communication in chimpanzees (Pan troglodytes)

DOHERTY, EMMA,ELIZABETH

A comparative developmental approach to multimodal communication in Chimpanzees (*Pan troglodytes*)

Emma Elizabeth Doherty

Thesis submitted in partial fulfilment of the degree of Doctor of Philosophy at Durham University

May 2023

Acknowledgements

"The whole of life is coming to terms with yourself and the natural world. Why are you here? How do you fit in? What's it all about?" - Sir David Attenborough

First and foremost, I thank my first supervisor, Zanna Clay for her guidance, wisdom and support. Both Zanna and my second supervisor Marina Davila-Ross helped me to formulate the proposal for this thesis while I was working in Borneo in the last volunteering position I could afford to do, and their belief in me will forever be appreciated. During my time as her PhD student, Zanna has been a great source of encouragement and her enthusiasm for my efforts has been able to bring me back from the depths of imposter syndrome many times. I have also been lucky to have a wonderful second supervisor, Marina, who has oftentimes challenged me to challenge myself which has helped me gain new perspectives and become an altogether more competent researcher.

I am grateful to have been a member of the C4D lab lead by Zanna. It has been great to be able to bounce ideas off one another and be a source of encouragement when needed. Thanks in particular to Jake Brooker and Georgia Sandars for making fieldwork in Zambia so enjoyable and for all their help in reliability coding. Jake has also been on hand to answer all my annoying statistics questions and I appreciate his unofficial position as the lab modelling guru.

I had a great experience conducting fieldwork at Chimfunshi Wildlife Orphanage, Zambia in 2021 and am thankful to Thalita Calvi and all the staff for being so accommodating and working to make the visit happen when the world was still upside down. It goes without saying that I am also grateful to the chimpanzees for tolerating being

followed around by a pale human with a weird accent for so many weeks. While I was there in the name of science, they provided so many funny and humbling moments, and the diversity in personalities reminded me why I wanted to do this work in the first place.

Of course, I am hugely thankful to my family and friends for a lifetime of support and encouragement, without who I would not have reached this point. Since I was a young child, my parents provided an environment in which my sisters and I believed we were capable of anything we worked hard at and actively fostered my enthusiasm for all things 'animals' when it became clear that was where my passion lay. To this day I am sure to get a text message or phone call to notify me of a new wildlife documentary I "would love" and dad still adds them to his Netflix watch list in case I ever happen to be around. My path to get here has involved grasping opportunities to gain new experiences in far-off places and I would not have had the courage and independence to do so if it wasn't for them. I am also thankful to my sisters Rebecca and Sarah, and dear friends Niamh, Nicole and Siún for always pretending to understand what it is I do and for all the conversations I left feeling motivated to go on. Lastly, I want to thank my cousin Amy, my biggest champion and built-in best friend for always being my shoulder to cry on and giving me back my confidence when I've lost it.

I would also like to extend many thanks to my examiners Professor Rob Barton and Professor Katie Slocombe for making my PhD examination such a pleasure and for their constructive feedback and insightful suggestions for the improvement of this thesis.

Lastly, I thank the British Association for Biological Anthropology and Osteoarchaeology, Lucie Burgers Foundation, the Durham psychology department and Faculty of Science for their generous financial contributions to my research.

Abstract

Studying how communication of our closest relatives, the great-apes, develops can inform our understanding of the socio-ecological drivers shaping language evolution. However, despite a now recognized ability of great apes to produce multimodal signal combinations, a key feature of human language, we lack knowledge about when or how this ability manifests throughout ontogeny. In this thesis, I aimed to address this issue by examining the development of multimodal signal combinations (also referred to as multimodal combinations) in chimpanzees. To establish an ontogenetic trajectory of combinatorial signalling, my first empirical study examined age and context related variation in the production of multimodal combinations in relation to unimodal signals. Results showed that older individuals used multimodal combinations at significantly higher frequencies than younger individuals although the unimodal signalling remained dominant. In addition, I found a strong influence of playful and aggressive contexts on multimodal communication, supporting previous suggestions that combinations function to disambiguate messages in high-stakes interactions. Subsequently, I looked at influences in the social environment which may contribute to patterns of communication development. I turned first to the mother-infant relationship which characterises early infancy before moving onto interactive behaviour in the wider social environment and the role of multimodal combinations in communicative interactions. Results indicate that mothers support the development of communicative signalling in their infants, transitioning from more action-based to signalling behaviours with infant age. Furthermore, mothers responded more to communicative signals than physical actions overall, which may help young chimpanzees develop effective communication skills. Within the wider community, I found that interacting with a wider number of individuals positively influenced multimodal combination production. Moreover, in contrast to the literature surrounding unimodal signals, these multimodal signals appeared highly contextually specific.

4

Finally, I found that within communicative interactions, young chimpanzees showed increasing awareness of recipient visual orientation with age, producing multimodal combinations most often when the holistic signal could be received. Moreover, multimodal combinations were more effective in soliciting recipient responses and satisfactory interactional outcomes irrespective of age. Overall, these findings highlight the relevance of studying ape communication development from a multimodal perspective and provide new evidence of developmental patterns that echo those seen in humans, while simultaneously highlighting important species differences. Multimodal communication development appears to be influenced by varying socio-environmental factors including the context and patterns of communicative interaction.

A note on the impact of Covid-19 on thesis research

My PhD research began in September 2019, with plans to begin data collection on immature chimpanzees living in large naturalistic enclosures in Chimfunshi Wildlife Orphanage, Zambia in Spring 2020. Of course, these plans were thwarted due to the global outbreak of Covid-19, which prevented me from travelling internationally or accessing chimpanzees at any alternative location. Fortunately, just prior to the outbreak, my second supervisor Dr Marina Davila-Ross shared with me part of her research groups extensive video repertoire of focal recordings of chimpanzees at Chimfunshi taken in the year 2017 and some from 2013. The original intention was to practice some coding protocols before travelling to Zambia, and to use in addition to any data I collected that year, expanding my dataset.

Thanks to this opportunity, during the spring/ summer of 2020 I was then able to fall back on the 2017 focal recordings in lieu of my own planned data collection, which contained follows of most members of three separate social groups including enough recordings of 11 age-appropriate individuals to extract communicative data. My original protocol required for a focal individual to be within 10 meters of at least one other individual, for a focal to commence. The 2017 video database typically provided annotations of individual identities present within 10 meters, allowing me to select particular videos that aligned with my own data collection protocols. Nevertheless, as this was not specifically part of their protocol, many focal recordings were not useable.

In 2021, I was able to return to Chimfunshi to complete planned data collection. This was valuable given my relatively uneven sample in terms of age and sex of focal individuals, as well as only being able to utilize a portion of the 2017 recordings. During

this time, I could collect new focal recordings of an additional 17 individuals. Although largely successful, I was unable to collect data for approximately 3 weeks of the 3-month period due to Covid-19 cases and subsequent isolations at the sanctuary.

Due to these delays and disruptions, completion of coding and analysis were also delayed. Originally, we had also intended on trying to collect parallel data on bonobos at Lola Ya Bonobo Sanctuary in D R Congo. Due to the pandemic delays and impact, this trip could not happen. Therefore, although the entirety of this thesis is concentrated on chimpanzees, I acknowledge the value in including both *Pan* species in this kind of research to assess our ancestral and derived communicative abilities with increased accuracy.

Table of contents

List of Tables

List of Figures

Figure 1.1 Video still showing a multimodal signal combination by an 11-year-old male (Little Jones). A 'bared teeth' facial expression and a 'reach' gesture were accompanied by a 'squeak' vocalization after receiving food-related aggression from an adult female. Photo taken by Emma Doherty at Chimfunshi Wildlife Orphanage, Zambia.

Figure 2.1. Effect of individual age (years) on the frequency of multimodal (MM) combination production in immature chimpanzees ($N = 28$). The mean number of MM combinations produced by each focal individual is shown. Area of the dots reflects the variation in sample size for each individual for each year of age. The solid line and dashed lines represent the fitted GLMM and 95% confidence intervals, respectively.

Figure 2.2. Effect of communicative context on the frequency of (a) unimodal (UM) signals and (b) multimodal (MM) combination production in immature chimpanzees ($N = 28$). Indicated are signal cases (dots), median (horizontal lines), quartiles (boxes), percentiles (2.5 and 97.5%, vertical lines) and distribution of data (shaded areas). Asterisks represent significance (* = <0.05; ** = <0.01; *** = <0.001).

Figure 2.3. Effect of individual age (in years) on the proportion of multimodal (MM) combinations produced by immature chimpanzees ($N = 28$). The proportional production of MM combinations in relation to unimodal signals is shown for each focal individual. Area of the dots reflects the variation in sample size for each individual for each year of age. The solid line and dashed lines represent the fitted GLMM and 95% confidence intervals, respectively.

Figure 2.4. Effect of communicative context on the proportional production of multimodal (MM) combinations relative to unimodal signals in immature chimpanzees ($N = 28$).

14

Indicated are signal cases (dots), median (horizontal lines), quartiles (boxes), percentiles (2.5 and 97.5%, vertical lines) and distribution of data (shaded areas). Asterisks represent significance (* = <0.05; ** = <0.01; *** = <0.001).

Figure 2.5. Effect of communicative context on the frequency of (a) facial and (b) gestural unimodal (UM) signal production in immature chimpanzees (N =28). Box plots show the mean number per context (open circles), median (horizontal lines), quartiles (boxes), percentiles (2.5 and 97.5%, vertical lines) and outliers (dots). Asterisks represent significance (* = <0.05; ** = <0.01; *** = <0.001).

Figure 2.6. Variation in mean number of multimodal (MM) combinations produced across age categories of immature chimpanzees ($N = 28$). Error bars represent mean $\pm$ 1 SD. F= facial expression; G= gesture; V=vocalization.

Figure 3.1. Effect of infant age (between) on the proportion of interactive behaviours mothers directed towards their infants that were signals compared to physical actions. The size of the dots reflects the variation in sample size of each individual. The solid line and dashed lines represent the fitted GLMM and 95% confidence intervals GLMM and 95% confidence intervals, respectively.

Figure 3.2. Effect of mean-centred (within) age on the proportion of maternal responses to infants' interactional behaviours (physical actions or signals). Raw proportions are presented separately for each centred month of age of a given individual; thus, individuals can be represented more than once. The area of the dots corresponds to the sample size per centred month of age. The solid line and dashed lines represent the fitted GLMM and 95% confidence intervals GLMM and 95% confidence intervals, respectively.

Figure 3.3. Proportion of infant interactive behaviour (physical actions vs signals) responded to by mothers. Indicated are individual observations (dots), distribution of data (shaded area), median (horizontal lines), quartiles (boxes), and percentiles (2.5 and 97.5%, vertical lines) across each behaviour type.

Figure 4.1. Effect of mean-centred number of interaction partners on the frequency of multimodal signal production in immature chimpanzees. Depicted is the mean frequency of multimodal combinations per communicative interaction across individuals (dots). The area of the dots corresponds to the sample size per individual. The solid line and dashed lines represent the fitted GLMM and 95% confidence intervals GLMM and 95% confidence intervals, respectively.

Figure 4.2. Effect of number of interaction partners on individual multimodal repertoire size of immature chimpanzees. Depicted is the multimodal repertoire size per individual (dots). The area of the dots corresponds to the sample size per individual. The solid line and dashed lines represent the fitted GLM and 95% confidence intervals, respectively.

Figure 4.3. Contextual distribution of multimodal repertoire in immature chimpanzees (a) number of combination types as a function of the number of contexts in which they were recorded (b) number of combination types recorded in each context.

Figure 5.1. Effect of signaller visibility and signaller age on probability of a signal event containing a multimodal signal. Depicted are the mean number of multimodal events produced for each year of age. The size of the dots corresponds to the sample size per individual. The solid line and shaded area represent the fitted GLMM and 95% confidence intervals, respectively.

Figure 5.2. Effect of multimodal combinations on the probability of eliciting a recipient response. Indicated are communicative interactions (dots) the mean number per condition (white circle), median (horizontal lines), quartiles (boxes), percentiles (2.5 and 97.5%, vertical lines). Asterisks represent significant values (*=<0.05; **=<0.01; ***=<0.001).

Figure 5.3. The effect of multimodal combinations on response latency [milliseconds] as a function of age [year]. Depicted are the mean response times for each year of age when a multimodal signal is absent or present. The size of the dots corresponds to the sample size per individual. The solid line and shaded area represent the fitted GLMM and 95% confidence intervals, respectively.

Figure 5.4. Effect of multimodal combinations on the probability of achieving an apparent communicative goal. Indicated are signal events (dots), median (horizontal lines), quartiles (boxes), percentiles (2.5 and 97.5%, vertical lines) and distribution of data (shaded areas). (*=<0.05; **=<0.01; ***=<0.001).

Chapter 1 - Introduction

Overview

In this introductory chapter, I will first review the topic of multimodal communication in (non-human) great apes with special focus on my study species, the chimpanzee *(Pan troglodytes)*. Then, I will introduce the primary topic of my PhD research: a comparative-developmental approach to the evolution of multimodal communication. As I will discuss further, the current research investigates addresses for the first-time developmental patterns of multimodal communication outside of humans. Chimpanzees were primarily chosen as the model species for this research first because of their close phylogenetic closeness with our species and resulting similarities in our extended developmental periods (Bründl et al., 2020) and complex social relationships that are navigated by means of diverse signalling behaviours (Goodall, 1986; de Waal, 1988).Variation in developmental patterns can then be used as a foundation from which to formulate approaches to this topic elsewhere in the primate order and beyond. Secondly, the bulk of comparative communication research in great apes concerning the use of different communicative behaviours (e.g., gestures, vocalizations), including recent multimodal signalling in adult apes, has been concentrated on chimpanzees (Liebal et al., 2014). While in general it will be important to consider how other primate taxa fit in the evolutionary development of complex multimodality in the future, the ample literature on chimpanzee communication provides a solid foundation from which to build a theoretical framework and to aid in interpretation of novel findings.

To begin, I will discuss the definition of multimodal communication, something which has varied across different research fields. From there, I will clarify the terminology I adopt in the study of multimodality and introduce this form of communication in the great-

apes. Although there has been a historical focus on unimodal signalling behaviours (i.e., gestures, vocalizations and facial expressions), there has also been a recent increase in multimodal perspectives, which has brought new evidence and questions regarding the prevalence, function and broader evolutionary relevance of multimodality, both in terms of its proposed communicative complexity and relevance to language evolution.

However, to truly understand how multimodal communication emerges in communicative interactions, an ontogenetic approach is critical. I will discuss this further, while also highlighting the dearth of developmental literature available on multimodality. In this respect, I review current understanding of signal development in apes, which at present is limited to findings from exclusively unimodal studies. I argue that developmental research, which includes how such signals are used in combination, is needed to understand the cognitive mechanisms underlying their production, and the impact of the socio-ecological environment. In turn, such investigation can shed light on the selective pressures which may have encouraged the evolution of combinatorial signalling in the hominid lineage which has culminated in human language.

What is multimodal communication?

Multimodal communication plays a key role in day-to-day human interpersonal interactions. Simultaneous control of acoustic and visual components (e.g., facial expressions, gestures, and body postures) of linguistic expression are consciously or unconsciously used to refine or emphasize specific messages during face-to-face interactions. So intricate is this inter-connectivity between speech and human bodily expressions, that researchers have suggested that an integrated neural process is involved in their simultaneous production (Marstaller & Burianová, 2015), sharing the same semantic meaning across different communication

channels (McNeill, 1992; Cassell et al., 1994; Vatikiotis-Bateson et al., 1996; Valbonesi et al., 2002; Kendon, 2004). From a perceiver's perspective, multimodality also matters greatly. Seminal studies have demonstrated that acoustic perception of human speech is modified via accompanying facial articulation (e.g., the McGurk effect; McGurk & MacDonald, 1976) and that humans rely on both visual and audio channels to understand a signal but giving more weight to the channel with the most reliable information (Massaro & Egan, 1996; Massaro, 1998).

Language has long been considered the defining characteristic of our species, divider between 'us' and 'them' regarding the rest of the animal kingdom. Consequently, the quest to understand how such a complex ability evolved has intrigued researchers for over a century and continues to be a topic of great interest today. While we have known for a considerable time that other animal species also communicate via signals simultaneously containing multiple sensory elements (e.g., olfactory, visual, auditory) (e.g., Darwin, 1872), research concerning the combined production of distinct communicative acts (e.g., vocalizations, gestures, facial expressions) is comparatively lacking. Instead, comparative research, which has typically focused on great apes (due to our close phylogenetic relationship), has investigated communicative acts in isolation rather than how they interact (Slocombe et al., 2011; Liebal et al., 2014). Its only more recently that comparative researchers have reached consensus that continuing to study different communicative behaviours separately is insufficient to capture the true complexity of primate communicative systems (see for review: Fröhlich and van Schaik, 2018).

From an evolutionary perspective, natural selection will favour individuals with the ability to use certain communicative tactics to influence a recipient's behaviour in a way which benefits them (e.g., Dawkins & Krebs, 1978). The simultaneous production of distinct

communicative behaviours is therefore of particular interest to understand how multimodality within language evolved as it has been argued that independent cognitive processes underlie the production of each (Waller et al., 2013), upon which selection can act. Moreover, studying multimodality from a production standpoint can shed light on external environmental factors influencing variation in signalling behaviour at the individual (e.g., context of production) and population level (e.g., social structure). Despite the value that studying multimodal signal production has for piecing together the evolutionary origins of complex signalling, this research is still in its infancy. Active investigation within this topic is needed to understand both the proximate (e.g., cognitive) and ultimate (i.e., adaptive) function of combinatorial signalling, which in turn will elucidate the degree of evolutionary continuity between non- human to human multimodal communication.

Competing terminology in multimodal research

Despite growing interest in the comparative study of multimodal communication, there still remains no true agreement regarding the definition of a 'multimodal' signal. Definitions tend to depend on the field of research and/ or the question being asked, as discussed below. While there is some crossover, in general seminal studies of multimodal communication can be divided into two main perspectives: those interested in the function of multimodal communication in terms of how it is *perceived* and those interested in the cognitive mechanisms underlying *production* (Higham & Hebets, 2013; Fröhlich and van Schaik, 2018).

In the field of behavioural ecology, researchers are primarily interested in evolutionary adaptation of animal behaviour in response to ecological pressures. When studying animal communication, behavioural ecologists therefore typically take the *perception* perspective, focussed on understanding the function of animal signals as they relate to the production

environment. In this kind of receiver-centric view, signal 'modalities' are classified based upon the sensory channel through which they are perceived e.g., visual, auditory, or tactile (Rowe, 1999; Parten & Marler, 1999, 2005). Function is then deduced by relating the sensory modality of a given signal to the circumstances of its production e.g., the addition of an auditory component to increase detection in a visually dense environment. Thus 'multimodal' signals in this field refer to combinations of more than one *sensory* modality, and function is then investigated via examining receiver responses to signals containing one vs multiple modalities (Smith & Evans, 2013).

Multimodal communication is also of interest to researchers focused on the evolutionary origins of human language. Comparative psychologists of ape communication may study this to gain perspective on the cognitive features and adaptive significance of behaviour, which can help piece together the selective drivers of shared features in our communication systems. Therefore, in contrast to the *perception* perspective taken in behavioural ecology, comparative research takes primarily a *production* perspective, in this case focusing on the proximate mechanisms underlying the production of distinct communicative acts e.g., gestures, vocalizations or facial expressions (Fröhlich & Hobaiter, 2018; Fröhlich & van Schaik, 2018). In this case, a multimodal signal refers to the specific combination of different communicative acts, regardless of the sensory channels through which they are perceived. For example, while a facial expression + a silent-visual gesture would be considered unimodal (visual) by behavioural ecologists, comparative psychologists classify this as a multimodal, to account for the fact that different cognitive mechanisms may operate, and interact, to produce a signal combination (Waller et al., 2013).

The alternative approaches to classifying multimodal communication, while each providing important contributions to our understanding of communicative complexity in other

species, have led to a tangle of terminologies making comparisons across fields and taxa problematic. Moving forward, it is important that clear, consistent definitions are established to improve comparability across research fields. This topic was recently reviewed at length by Fröhlich et al., (2019) who advocate for distinct terminology relating to the perspective taken on multimodal research moving forward. In receiver-centric research, a '*multimodal signal*' is then defined as signals consisting of two or more components of different sensory modalities which are obligatorily coupled (i.e., fixed) (e.g., lip smacking with a visual and auditory component); and in producer-centric research, a '*multimodal signal combination*' consists of the flexible combination (i.e. free) of two or more distinct signals whose components may be produced separately or combined (e.g., gesture plus vocalization) (Fröhlich & Hobaiter, 2018; Fröhlich & van Schaik, 2018). While I acknowledge the valuable contributions of perception-based approaches to the study and modelling of multimodal communication (e.g., Smith & Evans, 2013), given the focus of this thesis on the role of development in the production of communicative behaviour, I will thus investigate multimodality in reference to the combination of distinct communicative modalities. While perception-based perspectives offer valuable insights into the reception and decoding of multimodal signals (see p173 in Chapter 6 for further discussion). It is not the perspective upon which the following empirical research is based. Henceforth, I therefore adopt the latter term '*multimodal signal combination'* (sometimes shortened to multimodal combination) and it is this perspective which is discussed in the coming sections of this introduction.

Great ape multimodal communication

Much of the search for answers regarding the evolution of language has focused on great apes. As noted, this is primarily due to our close phylogenetic relationship and the resulting

insights they can provide to our common ancestral abilities. Yet, the traditional separation of focus on distinct behaviours (e.g., vocalizations, gestures and to a lesser extent, facial expressions) (Slocombe et al., 2011; Liebal et al., 2014) is also characterized by differences in approach, methodologies, and interpretations of signalling behaviour. This can make comparisons across modalities difficult and therefore restrict our ability to capture true communicative complexity (Partan & Marler, 1999). However, over the past decade, calls to unite the historically divided literature has led to progressive incorporation of multimodal perspectives to comparative research, revealing that like us, our closest living relatives also use multimodal signal combinations as part of their everyday communicative repertoires (e.g., Pollick & de Waal, 2008; Leavens et al., 2010; Clay et al., 2015; Genty et al., 2015, 2019; Taglialatela et al., 2015; Fröhlich et al., 2016b; Hobaiter et al., 2017; Wilke et al., 2017; see Figure 1.1).
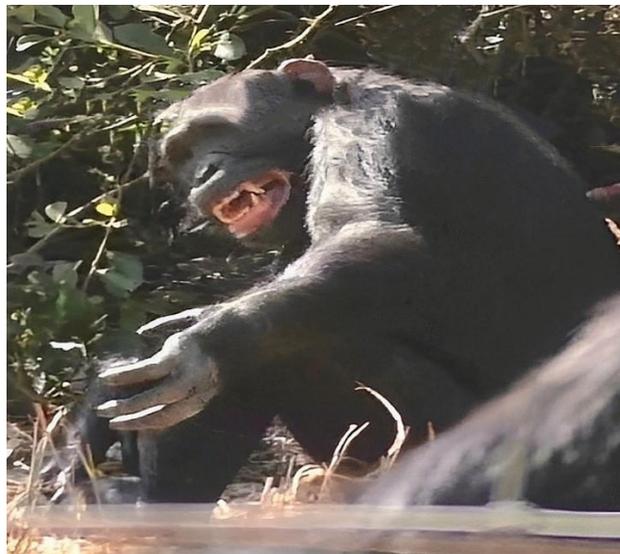
Figure 1.1 Video still showing a multimodal signal combination by an 11-year-old male (Little Jones). A 'bared teeth' facial expression and a 'reach' gesture were accompanied by a 'squeak' vocalization after receiving food-related aggression from an adult female.

Photo taken by Emma Doherty at Chimfunshi Wildlife Orphanage, Zambia.

As the study of multimodal communication in great apes is relatively new, we do not yet know the full extent to which it is manifested within the communicative systems of related species, nor how rich multimodal repertoires may be. Still, recent research has demonstrated that it is more common than previously thought. For example, in a study of captive chimpanzees, Taglialatela et al. (2015) reported that as much as 50% of vocalizations were accompanied by gestures, facial expressions (i.e., fear-grimace) or additional behaviours (e.g., chasing). More recently Wilke et al., (2017) observed a repertoire of 48 different multimodal signal combinations (gestures, vocalizations and facial expressions) among adults in a group of wild chimpanzees. Notably, in the latter study, although observed in 22 out of 26 focal individuals, multimodal combinations were produced at a significantly lower rate than unimodal signals.

This finding, that unimodal signals may be more dominant in the repertoire of adult chimpanzees, points to a potentially more specific function of multimodal combinations to disambiguate signaller's messages. This is supported by recent work on chimpanzees and bonobos (*Pan paniscus*) which have independently presented evidence that multimodal combinations add clarity to otherwise ambiguous signals (e.g., Wilke et al., 2017; Genty et al., 2015; Genty, 2019). For example, Hobaiter et al., (2017) showed that chimpanzees switch to vocal-gesture combinations when initial vocal signals were unsuccessful in reaching perceived communicative goals (but not when initial signals were gestures). Additionally, Genty et al. (2014) also showed that bonobos use the 'contest-hoot' vocalization in playful and aggressive contexts but add gestures to distinguish their intentions between the two. Refinement of a particular vocal signal via combination with

additional acts has also been shown in chimpanzees, where the function of a chimpanzee greeting call (pant grunt) was modulated by the addition of postures and gestures associated with submission when the approaching individual was higher ranking than the signaller (Fedurek et al., 2021). The latter two studies indicate that context is of particular importance in the production of multimodal combinations, particularly regarding interactions associated with higher costs of ambiguity, i.e., aggression.

Most of these recent studies have focused on the combination of ape vocalizations and gestures; however, the incorporation of facial expressions in studies of great ape multimodal combinations is currently lacking. In wild populations, this is likely due to difficulty in reliably observing and identifying them in their dense natural habitat. In addition, the historical characterization of facial expressions as 'emotional' signals linked to specific motivational states may have contributed to their exclusion from consideration as cognitively driven signals (Waller et al., 2017; Heesen et al., 2022). As such, facial expressions have more often been studied in the context of multimodal communication as part of '*multimodal signals'* (i.e., visual components), where there is evidence that chimpanzees cross-modally match facial expressions with corresponding vocalizations (Izumi & Parr, 2004; Parr, 2004).

However, recent research from a number of different ape species (including small apes) has provided evidence that they have at least some control over the production of different facial expressions (hylobatids: Schneider et al., 2016; orangutans *(Pongo pygmaeus):* Waller et al., 2015; chimpanzees (Davila-Ross et al., 2015; Palagi et al., 2019). Whether intentionally produced or not, to date we still have little understanding of the role of facial expressions in multimodal signal combinations. Apparently, the only study thus far

to address this is by Oña et al. (2019), who showed that in semi-wild chimpanzees, the addition of a particular facial expression to a gesture altered the likelihood of a specific response. When combined with a 'stretched arm' gesture, the 'bared teeth' face improved recipients' affiliative behaviour; however, it did not when combined with a 'bent arm' gesture. This was taken to suggest that while different facial and gestural components do not carry specific meanings, nor are meanings created consistently through combinations, facial expressions can also act to modify gestures, a property the authors referred to as 'componentiality.'

Overall, multimodal signalling is a clear feature of ape communication behaviour that has an apparent function in helping to disambiguate messages. While research on adult apes has demonstrated that the ability to flexibly combine distinct communicative acts into multimodal messages was present prior to the *Pan-Homo* split, what remains unclear is how they proximately arise within communication systems across the lifespan. In humans, the ontogenetic emergences of language must depend on prior cognitive capacities which have been shaped throughout complex developmental processes and experiences (Oller, 2000; Locke & Bogin, 2006). Thus far, comparative communication research in apes has focused almost exclusively on adults (Bard & Leavens, 2014); by comparison, we have little understanding of the cognitive mechanisms underlying communicative acts in other ape species, nor the role of ontogeny in shaping communicative outcomes.

Why study communicative development?

In evolutionary biology, it is now widely recognized that natural selection not only shapes phenotypic endpoints, but also the developmental processes resulting in those endpoints (Müller, 2007). By modifying developmental processes, natural selection can also produce novel

27

phenotypes by enhancing the complexity of pre-existing structures (Brigandt & Love, 2010).

Therefore, for comparative research to elucidate the evolutionary origins of multimodality, and

by extension- language, a developmental approach is critical. Understanding how features of

great ape communication systems develop through ontogeny will shed light on the cognitive and

functional mechanisms underlying signal production (Partan, 2013), which in turn will inform

our understanding of the external pressures shaping communication systems across evolutionary

time.

In humans, the importance of early social experiences in shaping socio-cognitive

outcomes is well established. Variation in social environment can affect many features

of human social, emotional, and interactional development (Maggi et al., 2010; Pillas et

al., 2014; Lopez et al. 2021), while lack of social stimuli can have extreme negative

impacts (e.g., Rutter et al., 2007). Extension of such considerations regarding

developmental experience have not, however, been applied as thoroughly in the study of

cognition in other ape species. Along with humans, non-human primates, particularly

the great apes, also share a longer period of development than is typical across the

animal kingdom (Bogin, 1990). Given the extended scope for learning and

environmental influence that comes with protracted early life-histories, developmental

research which focuses on early social experiences is vital to draw conclusions

regarding a species' socio-cognitive abilities. Indeed, studies where developmental

experience is considered in the examination of different socio-cognitive skills in

chimpanzee infants have drawn parallels between human and chimpanzee development,

in terms of the impact of variation in early interactional experiences and rearing

conditions (Bard & Leavens, 2014).

There is an extensive literature focused on the importance of early caregiver-infant interactions on human communication development. In humans, mothers and other caregivers are known play an active role in scaffolding infant communicative development via adjusting their communicative behaviour to their infant's developmental stage (e.g., Murphy & Messer, 1977; Schaffer et al., 1983; Adamson & Bakeman, 1984) and through prompt, appropriate and contingent responses to infant communicative attempts i.e., 'responsiveness' (Ainsworth et al., 1978; Bornstein & Tamis-LeMonda, 1989). In non-human apes, the first and primary interaction partner in a young individual's life is also their mother. In the first two years of life (van Lawick-Goodall, 1986), chimpanzees are in near constant physical contact with the mother, before an increase in spatial independence when they spend longer periods at greater distances (van Lawick-Goodall, 1968). As a consequence, the ape mother likely has an even greater influence on infants' early communicative development in apes than in humans, who by comparison tend to show cooperative caregiving (Hrdy, 2009) which entails the infant interacting with multiple caregivers. Thus, it is even more surprising that the role of ape mothers in early communicative development has so rarely been considered (but see Bard, 1992).

Changes in maternal signalling behaviour have been previously attributed to infant chimpanzees leaving the relative security of their mother's immediate proximity with age (Fröhlich et al., 2016). However, to my knowledge no study has investigated adjustments in maternal behaviour prior to this developmental milestone. Given what we know about the critical role mothers play in the communicative development of human infants, research focusing on this relationship in one of our closest relatives is needed to assess the degree to which ape mothers play a similar role in scaffolding communication behaviour. This could in turn have

wider implications for our understanding of early rearing on socio- communicative outcomes in apes more generally and thus for consideration of future study designs and interpretation of variation in adult behaviours.

Following this, although studies of earlier communication development are lacking, in general it has been suggested that meaningful communicative signals do develop first in interactions with the mother, before subsequently being shaped in interactions with other members of the social community (Plooij, 1978; Maestripieri & Call, 1996). Communication often occurs during complex social interactions, requiring skills which are dependent on a combination of cognitive, physical, and social development. However, although we know that social experiences strongly influence socio-cognitive outcomes (for reviews see Bard & Leavens, 2009; Leavens & Bard, 2011) and communicative development (Snowdon & Hausberger, 1997), this vital aspect has been rarely studied in great apes (Bard & Leavens, 2014).

Captive studies of chimpanzees have revealed that rearing environment has significant impact on the development of communicative abilities (Russell et al., 2011). Specifically, Russell and colleagues found that encultured apes (chimpanzees and bonobos) raised in human-like socio-communicative environments performed better in specific communicative comprehension and production tasks involving eye gaze and gestures than zoo-housed or laboratory raised individuals. While this does not necessarily represent an ecologically valid view of early communicative environments in these species, it does demonstrate that inherent communicative abilities can be shaped, or in this case leveraged, depending on early social input. More recently, Fröhlich and colleagues (2017) considered the influence of social factors on gestural production in a sample of wild chimpanzee infants. They found that gestural frequency and repertoire size were positively affected by

higher interaction rates with non-maternal conspecifics and the number of previous

interaction partners particularly in feeding and play contexts, thus contributing much

needed insight into the key role of peer-interaction in signal development.

While the aforementioned research has been important in demonstrating the value of

developmental approaches in understanding the socio-ecological drivers or great ape

communicative behaviour, thus far it has also been limited to single communicative acts in

isolation, most commonly gestures (e.g., Tomasello et al., 1985, 1989; Arbib et al., 2014;

Bard et al., 2014, 2017; Fröhlich et al. 2016, 2017). To date, studies considering the

developmental trajectories and shaping of facial expressions and vocalizations via social

exposure have been more limited. As human language emerges from developmental

processes, to understand its evolutionary origins we must adopt ontogenetic perspectives

which also consider how these communicative systems develop both separately and

together. In the coming sections, I will present existing research concerning the

development of facial expressions, vocalizations, and gestures separately, as well as the

value of incorporating a multimodal approach to this topic.

Development of facial communication

Although research on facial communication research in great apes is now growing, we still

know very little about the developmental trajectory. Existing research has tracked *when*

facial expressions emerge, revealing that in apparent contrast to the rapid facial development

of monkeys, great ape facial communication develops more gradually in a pattern similar to

that seen in human infants (Chevalier-Skolnikoff, 1982; Bard, 2003, 2007). For example,

new-born chimpanzees smile during rapid eye movement (REM), which shows a subcortical

maturation process parallel with human babies (Mizuno et al., 2006). Later in development,

chimpanzees raised in a human-reared environment have been indicated to show analogous 'cry' and 'smile' faces within the first two weeks of life (Bard, 2003), followed by pout faces, 'mad' faces and facial expressions associated with laughter (i.e., play faces) within the first 37 days of life (Bard, 2007). Observations of play faces in wild infants describe them as appearing later, between 6-11 weeks (van Lawick-Goodall, 1968; Ploij, 1984), accompanied by laughter at 12 weeks (van Lawick-Goodall, 1968). The earlier emergence of recognizable facial expressions in Bard's (2003, 2007) descriptions of neonate chimpanzees could be influenced both by clearer observation due to absences of a biological mother but also a degree of imitation of human caregiver expressions (Bard, 1998, 2004, 2007; Myowa-Yamakoshi et al., 2004). Moving forward, it will be important to extend research of facial communication development across groups living in more natural settings to be able to fully track patterns of facial expression emergence.

Aside from some tentative descriptions, we do not yet know if or how socio-ecological environments shape facial expressions during ape ontogeny. Again, this is likely due to the difficulty in reliably capturing facial expressions in wild populations and associated observational barriers in natural settings. Based on the structural similarities between immature and adult facial expressions (e.g., play face) and the low variability in facial repertoires within and between populations, it has been suggested that facial expressions are genetically channelled (see for review: Liebal et al., 2019). There is, however, indirect evidence of maturational changes in facial expressions in chimpanzee infants based on captive studies examining other aspects of behavioural development. For example, in a study of food-transfer between chimpanzee mothers and infants, Ueno and Matsuzawa (2004) showed that starting at 15 months, infants began to incorporate grin- and

grimace- like facial expressions with whimper vocalizations when food begging. The authors linked this incorporation of facial expressions in food solicitation tactics to the development of more complex patterns of food sharing. Additionally, focusing on the production of play faces during infant play contexts, Ross et al (2014) found that 15-month-old infants matched play faces during social play more frequently than they did at 12 months, indicating a developmental progression of emotional communication skills. Although not the focus of these studies, both describe longitudinal changes in facial communication associated with specific social contexts which could indicate that facial communication is subject to a degree of feedback from repeated interactions.

Lastly, conclusions regarding variation within facial expressions has traditionally relied on broad-category descriptions, but the incorporation of specialized facial coding software may shed new light on structural variation within facial expressions across ape ontogeny. For example, recent work conducted by Lembeck (PhD thesis: 2015) used a modified version of the facial action coding system (FACs, Ekman & Friesen, 1978) to analyse the structural properties (action units) of play faces produced by mother and infant chimpanzees and bonobos (via chimpFACs, Vick et al., 2007). While infant play faces aligned with a prototypical play face structure (Parr et al., 2007), responsive play faces of both chimpanzee and bonobo mothers exhibited a greater amount of morphological variation. Findings from this study therefore indicate that structures within expressions can progress from prototypical to more complex expressions but important questions remain: How does this happen? Is social interaction a factor in determining facial expression during ontogeny? Considering these important, yet unanswered questions, an approach that considers how these signals are integrated with signals from other communication modes

would be beneficial.

Development of vocal communication

Due to the traditional bias of (mis)characterizing language as speech (e.g., Hockett, 1960; Lieberman & Mc Carthy, 1975; Hauser et al., 2002), comparative research on the vocal behaviour of primates has been a focus of interest to uncover its evolutionary origins. Although typically biased towards monkey species (Slocombe et al., 2011; Liebal et al., 2022) increasing vocal research in apes has revealed evidence of numerous phylogenetic precursors of linguistic features present in their vocal systems – this includes vocal control (Lameira & Schumaker, 2019), intentionality (e.g., Schel et al., 2013; Townsend et al., 2017; Graham et al. 2020), functional reference (e.g., Slocombe & Zuberbuhler 2005, 2007; Schel et al., 2013; Watson et al., 2015) adjustment to conspecifics attentional state (Crockford et al., 2012, 2017) and syntax (see for review: Zuberbühler, 2020). However, understanding the developmental process underlying these language-like features of ape communication have been more overlooked until recently (Eaton et al., 2018).

Like facial expressions, earlier work on vocal development in chimpanzees described the types (van Lawick-Goodall, 1968; Plooij, 1984) and emergence of different vocalizations during infancy in relation to humans (e.g., Kojima, 2001, Bard, 2003). In more recent years, more work has been dedicated to investigating ontogenetic changes in ape vocal communication, revealing both developmental patterns as they relate to human speech development, and the role of the social environment in shaping vocal behaviour across time. For example, in a recent study of wild chimpanzee infant and juvenile call production, Taylor et al. (2021) examined age-related changes in the size and structure of vocal repertoires. In human infants, a transition from fewer acoustically graded vocal units to an

increased repertoire of distinct vocal units characterizes vocal ontogeny, a process known as the *expansion stage* (Oller, 2012; Vihman, 2014). Taylor and colleagues investigated whether vocal ontogeny in chimpanzees followed this same pattern but found that unlike humans, with increasing age chimpanzees showed greater acoustic graduation within vocal units as opposed to acquiring new ones. That is, chimpanzees appear to undergo an ontogenetic shift in the acoustic characteristics of a stable vocal repertoire.

While this research showed differences in chimpanzee and human vocal ontogeny, recent research has also highlighted similarities in vocal abilities between our species. Affective decoupling i.e., the ability to produce a vocal unit under a variety of affective states has been identified as a crucial evolutionary precursor to language (Oller et al., 2013). In humans, this capacity is present in early ontogeny and as such, may be more foundational to human speech than other elements of language faculty, such as proto-syntax or vocal elaboration (Oller et al., 2013), being prerequisites to speech development in general. Evidence for affective decoupling in adult apes has been found previously (e.g., Clay et al., 2015, Oller et al., 2019) but recently, Dezecache and colleagues (2021) found evidence for this capacity in early chimpanzee development. They found that while the common call type 'whimper' appeared tightly bound to negative affect, infant 'grunts' were observed across a range of affective contexts that were deemed positive, neutral, and negative. Like Taylor and colleagues, here too they observed acoustic gradation within 'grunt' calls, but gradation was linked to the affective context in which the call was produced, highlighting the importance of considering socioecological variables e.g., communicative context on signalling behaviour at various stages of ontogeny. Importantly, evidence for affective decoupling in infant chimpanzee calls could also indicate a degree of vocal 'functional flexibility' present

in early ontogeny, a foundational vocal capability of language production in humans.

That said, while some call types, such as 'grunts' may show evidence of functional flexibility, they may nevertheless also show specific acoustic variants that appear to be more constrained to a given communicative function. Furthermore, in the aforementioned studies, no other communicative modality was considered, the addition of which could further modify message 'meaning'. A recent example of this was presented by Genty (2019), who examined the production of infant bonobo 'pout moan' calls (akin to chimpanzee whimpers) in requests directed towards their mother. While all pout moans were used to solicit something from the mother i.e., associated with the same (negative) affective context as they were described in Dezecache and colleagues' study, the specific function of the call, to request carry/contact, food sharing, nursing or reassurances was related to the addition of a particular gesture type. The functional specificity of vocal-gesture combinations also increased with age indicating that this could be a learned skill. Although this paper was restricted by a small sample size, it presents an exciting possibility that if multiple communicative acts are considered simultaneously in the coding process across different functional contexts, we may find new understanding of the developmental mechanisms underlying signal production in young apes and by extension, the selective variables which encouraged vocal-gestural combinatorial signalling in the hominid line.

Development of gestural communication

In comparison to facial expressions and vocalizations, considerable research has been dedicated to understanding the proximate mechanisms underlying the acquisition and production of ape gestures, including across ontogeny. This topic, including the evidence to support different theories of acquisition has been reviewed extensively elsewhere (e.g.,

Liebal et al., 2014, 2019; Pika & Fröhlich, 2018) and as such, here I will only give a brief

overview of the major competing theories and what each can contribute to our

understanding of communicative development in apes (Table 1.1).

In general, there are currently four main theories ascribed to gestural development in

the great apes, the validity of which are subject to ongoing debate. First, is that gestures are

acquired through some form of social learning (social transmission). Within this theory, the

mechanism suggested to underlie social learning which has received the most attention is

imitation (Call & Carpenter 2002). It postulates that if gestures are acquired by copying

others, gestural repertoires should show high uniformity within groups, but they should

differ between groups (e.g., gestures would be group specific in presence or function).

Evidence put forward for social learning (imitation) included experimental studies which

showed that "encultured" chimpanzees, bonobos and orangutans raised by human caregivers

were able to reproduce familiar actions (e.g., Tomasello et al., 1993; Custance et al., 1995;

Myowa-Yamakoshi & Matsuzawa, 1999; Myowa-Yamakoshi et al., 2004; Call, 2001). That

said, while capable of reproducing more familiar actions or outcomes (i.e., "simple

imitation"), authors have generally agreed that non-human apes show difficulty copying

more novel actions or outcomes (i.e., "complex imitation") compared to human children

(e.g., Subiaul, 2016).

Second, it has been suggested that ape gestures are predominately innate, resulting

from a process of genetic channelling (Hobaiter et al., 2011a; Genty et al. 2009). This

process is also referred to as genetic transmission or phylogenetic ritualization. This view is

held by the cited proponents, largely based on findings that gestural repertoires appear

species-typical and show minor variation among individuals within or between groups and

sites. In terms of development, even if gestural repertoires are genetically channelled, it has been suggested that social learning still plays a part in 'fine-tuning' initially large, redundant repertoires subset of effective, regularly used gesture types via repeated social interactions (referred to as 'repertoire tuning', Hobaiter & Byrne, 2011b; Byrne et al., 2017). The smaller adult repertoires should contain gestures similar in form to those of younger individuals, but they are more flexible in their usage, with meaning derived from the context in which they are produced.

Third, a mechanism termed ontogenetic ritualization posits that previously non-communicative behaviours are shaped into increasingly ritualized, communicative gestures in repeated interactions with conspecifics (Call & Tomasello, 2007; Tomasello, 2008). For example, infant chimpanzees regularly physically 'take' food from the hands and mouths of their mothers while they are feeding. With increasing age and repeated interactions, this behaviour may become ritualized to become hand or mouth begging, where signallers use their fingers or lips to touch or stroke the hands and mouths of feeding recipients. To align with this theory, young individuals should have increasing gestural repertoire sizes with increasing age, reaching asymptote in adulthood. Furthermore, as the outcome of ritualization may vary for each dyad, individual repertoires would be highly variable and include idiosyncratic gestures. In contrast to social learning, individual repertoires should not overlap within or between groups and sites.

Lastly, the recently revised theory of social negotiation (Fröhlich et al., 2016b) is an amendment to seminal descriptions of signal development first provided over 40 years ago (Plooij, 1978; 1984) and a response to contradicting evidence from other theories of gestural acquisition. To a degree, this theory overlaps with that of ontogenetic ritualization but

instead of gestures becoming ritualized from previously non-communicative actions, (Fröhlich et al., 2016b; Pika & Fröhlich, 2018) supporters argue that they were "full-blown" behaviours to begin with. Here, gestural structure is modified over time via repeated interactions and thus subject to a learning process which contradicts the view of gestures as innate (Genty et al., 2009; Hobatier & Byrne, 2011b; Byrne et al., 2017). That said, the distinction between ontogenetic ritualization and social negotiation can, however, be difficult to make, since it is still possible that some aspects of shaping might have been overlooked (at least with the methods that have been used so far), whereas "full-blown" behaviours are much more salient and thus easier to observe.

Table 1.1 A summary of the major theories of gestural acquisition in great apes

Theory	Mode of acquisition	Variation within sites	Variation between sites
Social learning (transmission) [a]	Learn gestural repertoire from others e.g., via imitation	Low	High
Genetic channelling [b]	An innate, initially large and redundant repertoire is increasingly 'fine-tuned' to a subset of effective, regularly used signals	Low	High
Ontogenetic [c] **ritualization**	Previously non-communicative behaviours (e.g., hitting during play) become ritualized via repeated interactions into communicative	High	High

	gestures (e.g., arm raise to initiate play)		
Social negotiation [d]	Perlocutionary acts become conventionalized via interactional experience within the social environment.	High	Low

a. Call & Carpenter, 2002. b. Genty et al., 2009; Hobaiter & Byrne, 2011b; Byrne et al., 2017 c. Call & Tomasello, 2007; Tomasello, 2008. d. Plooij, 1978, 1984; revised in Fröhlich et al., 2016b.

To summarize, evidence for the acquisition of gestures in apes has been inconsistent. While some authors have suggested that multiple mechanisms could be involved in gestural acquisition/ production across development (Bard et al., 2014; Gasser & Arbib, 2019), in a recent review of this topic Liebal et al., (2019) put forward that this apparent inconsistency could be instead the result of methodological differences between major research groups. They argue for instance that: different research settings, definitions of a communicative gesture, samples (i.e., cross-section vs longitudinal) and focus on different aspects of results, more likely contribute to differences in perspectives on gestural acquisition. Therefore, research groups remain relatively limited by their own specific perspective instead of also considering how the findings of others are contributing to a bigger picture. The integration of perspectives and acknowledgment that more than a single mechanism is involved in gestural communication development is an important step for a better understanding of how apes acquire their gestures.

The need for a multimodal approach to communicative development

As discussed above, research on communication development in non-human apes to date has had a clear, unimodal focus. Multimodal signal combinations may have a different developmental trajectory altogether than unimodal signals so investigating their developmental pathways alongside one another is important to bring clarity to the mechanisms involved in signal acquisition within the complex and importantly, multimodal communicative systems of our closest living relatives.

If we think again about the competing theories of gestural acquisition, each could have some truth and contribute pieces to a bigger picture if we consider communicative signals more holistically. For example, it could be that apes are born with a rich, innate, and species-typical signal repertoires but their componential use as part as multimodal signal combinations could be shaped via additional mechanisms such as learning through repeated social interactions during ontogeny. Evidence for this may lie in the recent findings of emerging multimodal research discussed previously. For example, while gestural researchers have argued that gestures are contextually flexible signals, with meaning extracted from the context in which they are produced (Roberts et al., 2012; Hobaiter & Byrne, 2014; Byrne et al., 2017), multimodal signal combinations have been shown to be linked to specific communicative functions with specific combinations appearing to refine, clarify and disambiguate messages (e.g., Genty et al., 2014, 2015; Hobaiter et al., 2017; Genty, 2019; Ona et al., 2019). The production of these complex signals may require the development of more advanced cognitive skills or sufficient periods of learning to implement them effectively. On the other hand, perhaps multimodal combinations are replaced by more efficient unimodal signals (e.g., Fröhlich et al., 2016) and only used in

interactions where added clarification is required. Given that we do not yet know if multimodal signals share similar developmental patterns to unimodal signals, the degree to which they are genetically channelled or socially learned remains unresolved.

To address this, we first need to establish patterns of multimodal signal combination emergence during ontogeny alongside unimodal signals to establish comparability in developmental trajectories. From here, we can investigate the impact of the socioecological environment on the production of multimodal combinations and their function across communicative contexts to compare with patterns seen in unimodal research. Only by considering whole signals, including all their communicative components, can we make accurate assessments of communicative ontogeny in other great ape species. Furthermore, studying the development of multimodal communication in species with whom we are closely related can provide valuable insight into the factors which encouraged the manifestation of multimodal communication, overtaking unimodal signals as the primary mode of communication within the hominin lineage.

The social environment of developing chimpanzees

Before moving on to outline my thesis, here I will give a brief overview of the social landscape in which young chimpanzees develop, to provide a backdrop for my research. Aspects of social environment and development which are directly related to my empirical chapters are also discussed in further detail in respective chapters. Chimpanzees are highly social primates living in mixed-sex social groups, called "unit groups" or "communities", of typically 50–80 individuals (Nishida, 1968; Goodall, 1986; Nishida et al., 1990). Daily group composition is not stable but instead, individuals forage and socialize in smaller mixed age/ sex subgroups or "parties" that typically contain less than ten individuals and are

scattered throughout their home range. Party composition changes frequently according to both activity (e.g., feeding, resting) and distribution of resources (Pepper et al., 1999). Individuals therefore stay in close proximity with some conspecifics from the wider community at infrequent intervals, often up to weeks apart. This variation in subgrouping and ranging behaviour is referred to as a "fission- fusion" social organization. Reciprocated social relationships are a key feature of the chimpanzee social system and variation in association patterns related to fission-fusion dynamics is marked by varying degrees of time and energy investment in repeated and reciprocated instances of association and interaction across individuals (Watts, 2006; Mitani, 2009).

Male chimpanzees typically live for their entire lives in their natal groups forming strong, long-lasting social relationships with other males which facilitates the alliance-building and cooperation required for navigating through a highly competitive, political and territorial landscape (Goodall, 1986; Mitani et al., 2000; Boesch, 2009; Nishida, 2011). In contrast, 50-90% of females generally emigrate to neighbouring communities when they reach reproductive age (Mitani et al., 2002) where they must forge completely new relationships with members of their new community. Female adult lives are more solitary, dominated by mothering, competition for resources, and avoiding tense interactions with adult males (Williams et al., 2002).

Chimpanzee infancy is characterized by approximately 2 years of near constant physical contact with their mother at which time they become increasingly spatially independent and begin to have more opportunity to interact with individuals in the wider community (van Lawick-Goodall, 1967; Pusey. 1990). They remain nutritionally dependent on their mothers until weaning between the ages of 3-5 years but remain behaviourally dependent until around 8 years (van Lawick-Goodall, 1968). Over the juvenile period and

into adolescence, individuals establish themselves as increasingly independent and functional social units capable of sex-specific adult behaviours such as male boundary patrols or female engagement with infants (Pusey, 1990). Frequent engagement in social play is common in infancy and the juvenile period but decreases into early adolescence when adult sex roles and associated interactive patterns (e.g., reciprocated grooming, sex-related behaviours, aggression) become increasingly common (Pusey, 1990).

Outline of thesis

In my first empirical chapter (Chapter 2), I examine patterns of multimodal signal combination production in relation to unimodal signals across a range of behavioural contexts and different stages of chimpanzee ontogeny, from infancy to early adolescence. My aim in doing this is to (1) provide the first developmental trajectory of multimodal combination signalling in chimpanzees which can then be used as a basis from which to explore various external factors contributing to multimodal behaviour in the communicative systems of chimpanzees in future research; and (2) highlight potential differences in the developmental trajectory of unimodal vs multimodal communication. Differences between unimodal and multimodal combination ontogeny would further underscore the importance of taking multimodal perspectives to comparative communication research before drawing conclusions about developmental mechanisms influencing signalling behaviour in great- apes.

In the empirical chapters that follow, I then aim to explore external factors shaping communicative development in chimpanzees while maintaining this multimodal perspective. In Chapter 3, I focus on the role of the mother as an infant chimpanzee's first and primary interaction partner. Specifically, I am interested in addressing the current gap in great -ape literature concerning if/ how chimpanzee mothers adjust their communicative

behaviour to the developmental stage of her infant in the period prior to spatial independence and the degree to which they respond to their infants' communicative attempts. Moving on from this, I then look at the role of the wider social environment in contributing to patterns of multimodal combination production (Chapter 4) which are uncovered in Chapter 1. Here I am interested in how variation in communicative interactive behaviour with conspecifics contributes to multimodal signalling across development as measured by the production frequency and repertoire size of multimodal combinations. Furthermore, based upon previous evidence that signal combinations are less contextually flexible than unimodal signals (e.g., Genty et al., 2014; Genty 2019; Ona et al., 2019) I examine the specificity of multimodal combination repertoires across different behavioural contexts and whether this varies with age representing social learning. Lastly, in my final empirical chapter (Chapter 5), I look at the proximate role of multimodal combinations *within* communicative interactions. If multimodal signal combinations play a clarifying role in adult communication, I want to explore when this emerges in communicative ontogeny, and what cognitive and social factors contribute to this.

Chapter 2 - Multimodal communication development in semiwild chimpanzees

Abstract

Human language is characterized by the integration of multiple signal modalities, including speech, facial and gestural signals. While language likely has deep evolutionary roots that are shared with some of our closest living relatives, studies of great ape communication have largely focused on each modality separately, thus hindering insights into the origins of its multimodal nature. Studying when multimodal signals emerge during great ape ontogeny can inform about both the proximate and ultimate mechanisms underlying their communication systems, shedding light on potential evolutionary continuity between humans and other apes. To this end, the current study investigated developmental patterns of multimodal signal production by 28 semiwild chimpanzees, *Pan troglodytes*, ranging in age from infancy to early adolescence. We examined the production of facial expressions, gestures and vocalizations across a range of behavioural contexts, both when produced separately and as part of multimodal signal combinations (henceforth multimodal). Overall, we found that while unimodal signals were produced consistently more often than multimodal combinations across all ages and contexts, the frequency of multimodal combinations increased significantly in older individuals and most within the aggression and play contexts, where the costs of signalling ambiguity may be higher. Furthermore, older individuals were more likely to produce a multimodal than a unimodal signal and, again, especially in aggressive contexts.

Variation in production of individual signal modalities across ages and contexts are also

presented and discussed. Overall, evidence that multimodality increases with age in

chimpanzees is consistent with patterns of developing communicative complexity in human

infancy, revealing apparent evolutionary continuity. Findings from this study contribute novel

insights into the evolution and development of multimodality and highlight the importance of

adopting a multimodal approach in the comparative study of primate communication.

Introduction

In human language, both speech and signs are consistently integrated with additional,

visual information contained within gaze, facial expressions, gestures and bodily postures

(Levinson & Holler, 2014), which act to disambiguate my day-to-day communicative

interactions. Understanding the evolutionary origins of such a complex communication

system has long intrigued researchers, leading scientists to explore the communication

systems of my closest living relatives, the great apes (Liebal et al., 2014). Comparative

research has been fruitful in exposing fundamental building blocks of language shared by

other great ape (henceforth ape) species (e.g., Levinson & Holler, 2014; Crockford et al.,

2015; van Schaik, 2016; Fitch, 2017; Townsend et al., 2017, 2020) as well as shedding

light on the cognitive foundations and selective pressures that have shaped human

communicative evolution. Most of this research has focused on single modalities, namely

vocalizations or gestures, and, to a lesser extent, facial expressions (Liebal et al., 2014).

However, to understand how human communication has evolved, it is important to

acknowledge its multimodal nature.

It is first important to clarify what I mean by multimodal communication.

Historically, two contrasting definitions of a signal 'modality' have been adopted,

depending on the perspective taken on communicative function, that is, in terms of the sensory channel through which a signal is perceived (e.g., auditory, visual, tactile) or on the physical production of communicative acts (e.g. facial expression, gestures, vocalizations; see for a review Fröhlich & van Schaik, 2018). In the present study, while we acknowledge both definitions, we focus on multimodal signals as the coordination of distinct communicative acts. This is because, in terms of signal production, it has been argued that different cognitive processes might underlie the production of each (Waller et al., 2013). For example, a single communicative act can be perceived via multiple sensory channels (e.g., chimpanzee buttress drumming contains both audio and visual information) while different acts can be perceived via the same channel (e.g., a silent-visual gesture and a facial expression). We further distinguish between 'fixed' multimodal signals and 'free' multimodal signal combinations. Fixed multimodal signals (Smith, 1977; Partan & Marler, 2005) are those that contain obligatorily coupled components due to the mechanics of signal production (e.g., the chimpanzee 'pant hoot' face necessarily accompanies the 'pant hoot' vocalization). By comparison, 'free' multimodal signal combinations (Partan & Marler, 2005) are those whose components can be produced separately or be flexibly combined with other modalities (e.g., a facial expression and a manual gesture). This distinction also accounts for the potential variation in cognitive processes that underlie their production.

Lastly, we acknowledge discrepancies in the reported identification of multimodal signal combinations. While fixed multimodal signals occur simultaneously by default, this has not been a requirement under definitions of multimodal signal combinations in the past (Pollick & de Waal, 2007). However, for consistency with more recent research (e.g., Luef & Pika, 2017; Wilke et al., 2017; Fröhlich et al. 2019; Oña et al., 2019), I consider

multimodal signal combinations (sometimes shortened to as multimodal combinations) as temporally co-occurring combinations of facial expressions, gestures and vocalizations.

As noted, previous investigations of communicative homology among apes have largely been unimodal, that is, focused on a single communicative act in isolation (e.g., gestures or vocalizations; Slocombe et al., 2011; Liebal et al., 2014). This makes comparisons across modalities difficult and restricts my ability to capture potential communicative complexity (Partan & Marler, 1999). That said, a recent increase in multimodal communication research has revealed that other ape species also produce rich, overlapping combinations of different signal modalities as part of their everyday repertoires (e.g., Pollick et al., 2008; Genty et al., 2014; Taglialatela., 2015; Fröhlich et al., 2016b; Wilke et al., 2017; Oña et al., 2019). In one study of captive chimpanzees, *Pan troglodytes*, as much as 50% of their vocalizations were accompanied by gestures and/ or facial expressions (Taglialatela et al., 2015). More recently, Wilke et al. (2017) recorded a total of 48 distinct multimodal combinations used by a population of wild chimpanzees, although unimodal signals were more frequently produced overall. While these studies reveal that multimodal communication is present in apes, knowledge of the degree of continuity between ape and human multimodality, as well as the ultimate drivers of my shared ability to combine multiple signal modalities, remain limited.

Developmental research is key to understanding the basis of communication, including its multimodality (Partan, 2013). Tracking the ontogenetic emergence of different signalling behaviours is needed to identify the proximate mechanisms underlying their production and, by extension, the effects of socioecological variation on communicative outcomes (see Bard & Leavens, 2014). Understanding the development of

multimodal communication in my closest primate relatives can have important

implications for my understanding of the evolutionary processes that have acted upon

complex signalling in the hominid lineage. Thus far, the ontogeny of primate multimodal

signalling has been studied in terms of sensory integration (Dafreville et al., 2021) and

cross-modal perception (see for a review Ghazanfar, 2013). By contrast, much less

attention has been dedicated to the coordination of distinct communicative acts. Different

mechanisms may underlie the production of different communicative behaviours,

regardless of the senses used to perceive them (Waller et al., 2013). Understanding the

cognitive processes involved in signal integration and how they are shaped throughout

development will shed light on the impact of the socioecological environment on

combination signalling and, thus, potential evolutionary continuity between ape and

human multimodality. Multimodal signal combinations may have a different

developmental trajectory to the production of unimodal signals, although little is yet

known about this in apes (Gillespie-Lynch et al., 2014; Liebal et al., 2014). It could be that

unimodal communication emerges first, as is the case in human illocutionary

communication, where infants typically develop increasing control over the coordination

of separate modalities before the emergence of synchronous gesture–speech combinations

(e.g. Bretherton & Bates, 1979; Acredolo & Goodwyn, 1988; Iverson & Thelen, 1999;

Iverson & Goldin-Meadow, 2005; Esteve-Gibert & Prieto, 2014; Murillo et al., 2018)

However, previous evidence from chimpanzees' points to a potentially different

developmental pattern of communication than is observed in humans. For example,

looking at unimodal serial gesturing in wild chimpanzees, Hobaiter and Byrne (2011b)

reported that adults used fewer gestures and gestural sequences than younger individuals.

With increasing age, individuals were more likely to use a smaller number of more effective gestures, a process the authors termed 'repertoire tuning'. In addition, while multimodal research is overall scarce, infant chimpanzees have been observed to produce multimodal combinations early in ontogeny (Bard et al., 2014) and their communication has been reported to move from more multimodal (vocal–gestural) combinations to unimodal (gestural) signals across development (Fröhlich et al., 2016b). Although these examples appear to indicate that chimpanzee communication may move from more multimodal to unimodal systems across development, current insights are limited. Not only is this research still rare, but it has either been exclusively unimodal (i.e., gestural) or restricted to one behavioural context.

More recently, research has also begun to demonstrate that, like humans, ape multimodal combinations function to aid comprehension by disambiguating or complementing a core message (Genty et al., 2014, 2015b; Hobaiter et al., 2017; Wilke et al., 2017; Genty, 2019; Oña et al., 2019). If this is the case, we would hypothesise that it is most important to produce unambiguous messages when the costs of misunderstandings are highest, such as under risk of physical aggression. Further evidence is needed to test this, which represents one goal of the present study. Furthermore, it is also possible that the relative advantage of communicative clarity according to context may vary as a function of developmental stage in chimpanzees. For example, steady increases in solid food intake (e.g., Bray et al., 2018) together with increased spatial and behavioural independence from their mothers (van Lawick-Goodall, 1968; Pusey, 1990) can expose older individuals to higher levels of feeding competition and associated aggression risks. Thus, in addition to the need to investigate the production of multimodal combinations in

apes, a more inclusive approach, capturing a range of behavioural contexts, is required to judge developmental patterns of multimodal communication more accurately.

In the current study, we addressed these points by investigating the effect of age and contextual factors on developmental patterns of multimodal communication in chimpanzees. Using a cross-sectional sample of semiwild immature chimpanzees, ranging in age from infancy to early adolescence, we analysed the production of commonly described facial expressions (Parr et al., 2005, 2007; Bard et al. 2011), gestures (Nishida et al., 1999; Byrne et al., 2017) and vocalizations (van Lawick-Goodall, 1968; Plooij, 1984; Kojima, 2008; Slocombe & Zuberbühler, 2010) that occurred both singularly (unimodal signals) and as part of multimodal combinations. We tested the following hypotheses and predictions. First, given their close phylogenetic relationship to humans we expected chimpanzees to share a comparable developmental pattern of increasing communicative complexity. That is, we predicted that the frequency and relative production of multimodal combinations would be higher in older individuals. Second, according to the hypothesis that multimodal combinations serve to disambiguate meaning in ape communication, we predicted that the frequency and proportional production of multimodal combinations would be highest in the context of aggression, where the costs of ambiguity may be particularly high. In addition, we expected this relationship to be stronger in older individuals given the potential increased risk of aggressive social encounters as individuals start to establish themselves more in their social network. Lastly, although not our main focus, we also provide a detailed repertoire of multimodal combinations produced by immature chimpanzees across age categories and contexts to help inform future work.

Methods

Study site and subjects

Data were collected at Chimfunshi Wildlife Orphanage Trust (hereafter Chimfunshi), a chimpanzee sanctuary located in the Copperbelt region of northern Zambia (12°23′S, 29°32′E). The Chimfunshi population comprises four socially stable groups which are accessible for non-invasive observations. Enclosures measure between 20 and 77 km^2 and primarily consist of miombo woodland, which offers chimpanzees the opportunity to exercise natural behaviours including foraging, climbing and nest building in species-typical fission–fusion systems (Ron & McGrew, 1988). Chimpanzees spend all day and night in the outdoor habitats. The possible exception is for 1–2 h in the middle of the day when keepers provide access to an indoor area for extra food provisioning. Any individual could enter or leave the indoor space during this brief time, although it was typically more dominant individuals that used this space. The majority of individuals remained outdoors, and no data were collected from indoors. Concurrent food provisioning occurs at enclosure fence lines, where keepers carefully distribute additional food items to all individuals remaining outdoors. All chimpanzees observed in the present study were born at Chimfunshi and have grown up within one of the three social groups included in my sample. Therefore, direct interaction with or handling by human carers has been minimal, allowing all behaviours to remain relatively species-typical and the ecological validity of my sample to remain higher than for some captive populations.

The average age of sexual maturity in a chimpanzee female is 11.5 years (Walker et al., 2018), and although males experience an earlier sexual adolescence (approximately 8–10 years) they do not tend to become socially independent until around 12 years of age

(Pusey, 1983, 1990). Therefore, the cut-off age of inclusion for both sexes in my study was 11.5 years because both sexes reach social independence at around this time. If an individual was <11.5 years at the beginning of the field period (May–September 2017; May–August 2021) but surpassed this before it ended, they were still included in the sample. As all individuals in this study were born at Chimfunshi, their dates of birth were known to at least the month and year. The final sample includes 28 individuals, with 10 infants, seven juveniles and 11 early adolescents (Table 2.1). Age classes were assigned following van Lawick-Goodall's (1968) age classifications: infant (0–4 years), juvenile (5–7 years) and early adolescent (8–11 years).

Table 2.1. Information on 28 observed individuals including age (in categories and years), sex, group and total number of hours observed.

Age (category)	Age (year)	No. of individuals (male/female)	Group (no. in group 1/2/4)	Observation time (h)
Infant	1	2 (1/1)	(1/1/0)	4.61
	2	2 (1/0)	(0/2/0)	4.57
	3	4 (3/1)	(2/2/0)	6.06
	4	2 (1/1)	(0/1/1)	11.71
Juvenile	5	5 (1/4)	(3/2/0)	13.52
	6	2 (1/1)	(0/1/1)	7.25
Early adolescent	8	1 (0/1)	(0/1/0)	1.49
	9	5 (3/2)	(4/1/0)	13.57
	10	3 (1/2)	(2/1/0)	12.79
	11	2 (0/2)	(1/1/0)	8.56

Data collection

Data were extracted from recorded focal observations carried out at enclosures 1, 2 and 4 between May and September 2017 and between May and August 2021 where group sizes ranged from 12 to 58 individuals (see Appendix 1.1 Table A1.1 for full group size information from each observation year). Focal observations were not made at enclosure 3 due to lack of sample-appropriate individuals and poor visual conditions. We used a focal- animal sampling approach (Altmann, 1974). Focal individuals were recorded for 5 min periods once or twice a day during morning (0730–1200) or afternoon (1200–1730) sessions. The focal individual was selected opportunistically but where multiple focal individuals were visible, priority was given to those with fewer observations at that point. Given our focus on communication, observations were only started when an individual was within 10 m of another individual and/or there was potential for a communicative interaction. We define a communicative interaction as an interaction between two individuals during which a signal was produced. A communicative interaction lasted from the interaction onset until one individual leaves, or there is a clear break in the interaction >10s as identified by break in attention between focal and recipient. In the case of communicative interactions involving continued active contact between individuals e.g., physical play, embracing, grooming; if signalling ceased but contact was maintained, this was still coded as a continuous interaction until contact was broken and there was no further contact or interaction >10s. Focal observations were recorded using an HD camcorder (Panasonic HC-VX870) with an external unidirectional microphone (Sennheiser MKE 400). A total of 84.13 h of observations were collected across 28 individuals (mean ± SD = 3.00 ± 1.61 h per individual).

Following definitions established previously in the study of chimpanzee

communication (Schneider et al., 2012; Hobaiter et al., 2017), we coded nine behavioural contexts (Table 2.2) that considered the information provided before and after a signal was produced. For example, if a signal was produced that was associated with the initiation of a behaviour, for example play or affiliative contact, the context was coded as such. We believe that, by doing so, I captured the most accurate representation of signalling context. However, signals were only produced commonly enough in three of these behavioural contexts to be examined further with regard to unimodal signal and multimodal combination production (feeding, play and aggression). To take full advantage of all recorded observations and establish a fuller picture of multimodal combination frequency and the types of signal combinations produced, we still included all contexts in my analyses. Remaining contexts were included in an 'other' context category which was used as the reference level in our statistical models. See Appendix 1.1 Table A1.2 for full reporting of the 'other' category.

Table 2.2. Description of behavioural contexts for communication in immature chimpanzees

Behavioural context	Description
Access [a, b]	Behaviours related to the access of an object such as offering or preventing access
Affiliation [a, b, c]	Behaviours with the apparent aim of decreasing distance or requesting physical contact. Includes unaggressive approaches and greeting events
Aggression [b, c]	Initiation of or response to aggressive behaviours including
Feeding [c]	Individuals engaged in behaviours related to food intake (e.g.,

Grooming [a, b, c]	Behaviour accompanying the request of or participation in
Rest [a, c]	Behaviour occurring when individuals are stationary without participation in physical activity (e.g., lying down)
Play (social) [b, c]	Two or more individuals engaging in playful behaviour
Sexual [a, b]	Behaviour accompanying sexual interaction, e.g., presenting
Travel [a, b]	Behaviour accompanying locomotion in the enclosure.

a Included in the 'other' context category. b Schneider et al. (2012). c Hobaiter et al. (2017).

Behavioural Coding

A total of 808 clearly visible communicative interactions were observed across 394 video recordings. For each communicative interaction, all observed occurrences of facial expressions, vocalizations and gestures produced by the focal individual were coded using ELAN (version 6.0) open-source video annotation software (https://archive.mpi.nl/tla/elan). Signals were assigned to one of three signal categories for analysis: unimodal (UM), multimodal (MM) combination or fixed MM signals. However, sample sizes of fixed MM signals with a clear communicative partner were too low to conduct inferential statistics. Therefore, only unimodal and multimodal combinations are discussed. In line with past unimodal research (Hobaiter & Byrne, 2011b; Genty et al., 2014; Graham et al., 2018), instances where single or multiple signal modalities were produced in quick sequence (i.e., <1 s pauses between signal units) were categorized differently and assigned to unanalysed categories. Full details of these unanalysed categories are provided in Appendix 1.2 Including definitions and justifications for exclusion. For information on the total number of signals coded within each signal category see Appendix 1.2 Table A1.3.

To enable greatest comparability and consistency with previous research, we relied

on established definitions of signal types across modalities in chimpanzees (types of facial expressions: Parr et al., 2005, 2007; Bard et al., 2011; vocalizations: van Lawick-Goodall, 1968; Plooij, 1984; Kojima, 2008; Slocombe & Zuberbühler, 2010; gestures: Nishida et al., 1999; Byrne et al., 2017). Signal types coded within each modality are given in Table 2.3. Six types of facial expressions were used in this analysis, based upon prototypical chimpanzee expressions (Parr et al., 2005; see Appendix 1.3, Table A1.4) which have been further validated via specific combinations of facial muscle movements using a chimpanzee Facial Action Coding System (i.e., chimpFACS, Vick et al., 2007; Parr et al., 2007; see also Bard et al., 2011). For the coding of vocalizations, we relied on eight broad categories of vocalizations known to be produced by young chimpanzees (see Taylor et al., 2021; see Appendix 1.3 Table A1.5). This helped ensure intercoder reliability (see below) as while there is general agreement regarding the call types produced by young chimpanzees (e.g., grunts), the extent to which young individuals produce distinct subtypes (e.g., food-grunt, pant-grunt, etc.) remains understudied (Taylor et al., 2021). A gesture is defined here as directed, nonlocomotory movement of the head, limbs, or body and where the signaller showed anticipation of a recipient's reaction via eye gaze and/or body orientation (Call & Tomasello, 2007). A total of 50 gestures were coded from my video footage based largely on the repertoire proposed by Byrne et al. 2017 (see Appendix 1.3 Table A1.6). For details of fixed multimodal and different multimodal combinations coded including the context in which they were produced, see Table 2.4.

Table 2.3. Facial, vocal and gestural signal types observed in immature chimpanzees (*N*=28)

Signal modality	Total	Signal types
Facial expressions [a]	6	Bared teeth face, open mouth face, pant hoot face, pout face, scream face, whimper face
Vocalizations [b]	8	Bark, grunt, huu-call, laughter, pant hoot, squeak, scream, whimper
Gestures [c]	50	Arm raise, arm swing, beckon, big loud scratch, bipedal stance, bite, bow, crouch, dangle, directed push, drum other, embrace, finger in mouth, gallop, grab, grab-pull, hand fling, hand on, head shake, head stand, hit with object, jump, kick, look, mouth stroke, object in mouth approach, object move, object shake, pirouette, poke, pounce, present body part, present genitals, punch ground, push, reach, roll over, rub rump, side roulade, slap object/ ground, slap other, smack lips, somersault, stomp, stomp other, tap other, touch other.

[a]Parr et al., 2005, 2007; Bard et al., 2011. [b] van Lawick-Goodall, 1968; Plooij, 1984; Kojima, 2008; Slocombe & Zuberbühler, 2010. [c]Nishida et al., 1999; Byrne et al., 2017.

Table 2.4. Observed repertoire of multimodal combinations in immature chimpanzees. Given are descriptions of overlapping behaviors, signal modalities within each combination, the total number of occurrences observed, the number of individuals and age class who produced the combination and the context in which they were recorded.

Combination	Modality	Cases	Age	Context
Arm raise + grunt	FG	1	1 J	Affiliation
Arm raise + whimper	GV	1	1 J	Feeding
Bared teeth + arm raise + squeak	FGV	1	1 J	Deeding
Bared teeth + bark	FV	1	1 EA	Aggression
Bared teeth + directed push + grunt	FGV	1	1 EA	Aggression
Bared teeth + embrace + squeak	FGV	1	1 J	Play
Bared teeth + grab + squeak	FGV	1	1 EA	Aggression
Bared teeth + hand fling	FG	2	1 J, 1 EA	Aggression
Bared teeth + hand fling + squeak	FGV	2	1 EA	Aggression
Bared teeth + present genitals	FG	1	1EA	Sex
Bared teeth + reach + whimper	FGV	3	1 I, 1 J, 1 EA	Feeding
Bared teeth + slap other	FG	1	1 EA	Aggression

Bared teeth + squeak	FV	25	4 J, 8 EA	Affiliation, aggression
Bared teeth + tap other + squeak	FGV	3	2 EA	Aggression
Bared teeth + touch other + squeak	FGV	2	1 EA	Aggression, feeding
Bared teeth + whimper	FV	1	1 I	Feeding
Bipedal rock + hoo	GV	1	1 J	Sex
Bipedal stance + slap other + grunt	GV	1	1 EA	Aggression
Bite + grunt	GV	2	1 I, 1 EA	Aggression, feeding
Bite + squeak	GV	2	1 J, 1 EA	Affiliation, aggression
Crouch + bark	GV	3	2 EA	Aggression
Crouch + grunt	GV	5	3 EA	Aggression
Crouch + pant hoo	GV	1	1 EA	Aggression
Crouch + tap other + bark	GV	2	2 EA	Aggression
Crouch + tap other + grunt	FG	1	1 EA	Aggression
Directed push + whimper	GV	2	1 I, 1 EA	Affiliation, feeding
Drum other + grunt	GV	1	1 J	Feeding
Gesture + whimper	GV	1	1 EA	Feeding
Grab + scream	GV	1	1 I	Aggression
Hand fling + grunt	GV	1	1 J	Play

Hand fling + scream	GV	1	1 EA	Aggression
Hand on + scream	GV	2	1 EA	Aggression
Hand on + squeak	GV	2	1 EA	Aggression
Hand on + whimper	GV	3	2 I, 1 J	Affiliation
Mouth beg + whimper	GV	3	1 J, 1 EA	Feeding
Object move + hand fling + grunt	GV	1	1 EA	Aggression
Open mouth + arm raise + grab	FG	1	1 J	Play Play
Open mouth + arm shake	FG	1	1 J	Play
Open mouth + arm swing	FG	2	1 J, 1 EA	Play
Open mouth + bipedal stance	FG	1	1 J	Play
Open mouth + bipedal stance + grab	FG	1	1 I	Play
Open mouth + bipedal stance + tap other	FG	1	1 I	Play
Open mouth + bow	FG	1	1 J	Play
Open mouth + climb on	FG	3	1 I, 1 J	Play
Open mouth + crouch + bark	FGV	1	1 EA	Aggression
Open mouth + dangle	FG	1	1 I	Play
Open mouth + drum other	FG	2	1 I, 1 J	Affiliation, play
Open mouth + grab	FG	3	1 I, 1 J	Play
Open mouth + grab-pull	FG	3	1 I, 1 J, 1 EA	Play

Open mouth + grunt	FV	1	1 J	Play
Open mouth + hand fling	FG	1	1 I	Play
Open mouth + hand fling + slap other	FG	1	1 EA	Aggression
Open mouth + head nod	FG	1	1 I	Play
Open mouth + hit other with object	FG	4	1 I, 1 J, 1 EA	Aggression, play
Open mouth + laugh	FV	78	8 I, 3 J, 3 EA	Play
Open mouth + laugh + grab other + slap other	FGV	1	1 EA	Play
Open mouth + object move	FG	2	1 I, 1 J	Play
Open mouth + pirouette + stomp	FG	1	1 J	Play
Open mouth + poke	FG	1	1 EA	Play
Open mouth + pounce	FGV	1	1 I	Play
Open mouth + reach	FG	1	1 I	Play
Open mouth + roll over	FG	12	4 I, 6 J, 1 EA	Play
Open mouth + slap other	FG	11	3 I, 3 J, 1 EA	Aggression, play
Open mouth + somersault	FG	16	3 I, 4 J, 1 EA	Play
Open mouth + somersault + laugh	FGV	1	1 EA	Play
Open mouth + somersault + touch other	FG	2	2 J	Play

Open mouth + stiff walk	FG	2	1 J, 1 EA	Play
Open mouth + stomp	FG	5	1 I, 3 J	Play
Open mouth + stomp + somersault	FG	1	1 J	Play
Open mouth + stomp + throw object	FG	1	1 J	Play
Open mouth + tap other	FG	7	1 I, 4 J, 1 EA	Play
Open mouth + bow + roll over	FG	1	1 I	Play
Open mouth + head stand	FG	1	1 I	Play
Pout + beckon	FG	1	1 J	Access
Pout + bipedal stance + present genitals	FG	1	1 EA	Affiliation
Pout + crouch + reach	FG	1	1 EA	Aggression
Pout + grunt	FV	2	1 J, 1 EA	Play, rest
Pout + hand fling	FG	3	1 I, 1 J, 1 EA	Access, play
Pout + hand on	FG	2	1 I	Access
Pout + object move	FG	1	1 EA	Rest
Pout + poke	FG	1	1 I	Grooming
Pout + reach	FG	4	1 J, 2 EA	Feeding, grooming, rest
Pout + reach + grunt	FGV	1	1 J	Affiliation
Pout + slap ground + grunt	FGV	1	1 J	Feeding

Pout + slap ground + stomp	FG	1	1 EA	Access
Pout + slap other + grunt	FGV	1	1 I	Aggresion
Pout + slap other + stomp	FG	1	1 J	Play
Pout + stomp	FG	2	2 EA	Aggression
Pout + stomp + bark	FGV	1	1 I	Aggression
Pout + tap other	FG	2	1 I, 1 J	Affiliation, grooming
Pout + tap other + grunt	FGV	1	1 J	Affiliation
Pout + touch other	FG	2	1 J	Aggression, feeding
Pout + touch other + grunt	FGV	3	2 J	Aggression
Present genitals + squeak	GV	1	1 EA	Feeding
Present genitals + whimper	GV	1	1 EA	Sex
Reach + grunt	GV	2	1 J, 1 EA	Feeding
Reach + scream	GV	1	1 RA	Aggression
Reach + whimper	GV	11	2 I, 2 J, 2 EA	Aggression, feeding, grooming, travel
Slap ground + pant hoot	GV	1	1 J	Rest
Slap ground + scream	GV	1	1 J	Aggression
Slap other + grunt	GV	1	1 EA	Play
Stiff walk + grunt	GV	1	1 EA	Aggression
Stomp + grunt	GV	1	1 EA	Aggression
Tap other + grunt	GV	2	1 I, 1 EA	Feeding, play
Tap other + scream	GV	1	1 I	Feeding

Tap other + squeak	GV	3	I J	Aggression
Touch other + bark	GV	1	1 EA	Affiliation
Touch other + grunt	FG	6	1 J, 3 EA	Affiliation, aggression
Touch other + scream	GV	1	1 EA	Feeding
Touch other + whimper	GV	8	3 I, 1 J, 1 EA	Feeding

Modality: F = facial expression, G = gesture, V = vocalization; Age: I = infant, J = juvenile, EA = early adolescent.

Inter-coder reliability

To assess the reliability of video coding, a second and third independent researcher also coded all signal events across 15% of the total number of video recordings (57 focal recordings, 27 individuals). Cohen's kappa (Cohen, 1960) was calculated for the reliable identification of each modality independently across coded signal events as well as when a signal event contained just one or multiple modalities simultaneously. The mean kappa value obtained for each modality indicated excellent levels of coder agreement (Fleiss, 1981; facial expressions = 0.83; vocalizations=0.92; gestures=0.81). The level of agreement regarding the singular or simultaneous production of signal modalities was also excellent (0.81).

Statistical Analyses

We used generalized linear mixed-effects models (GLMMs) to investigate developmental patterns of unimodal and multimodal combinations produced across our sample. For all models, we included age (in years; range 1–11) and context of signal production (feeding, play, aggression and 'other' (reference level)) as test predictors, while also controlling for

sex (male, female), group number (1, 2 and 4) and observation year (2017, 2021). Given that

the effect of context on signalling behaviour could change over the course of ontogeny, we

also included a two-way interaction term between age and context in all models. As a

random effect (intercept) we included 'Signaller ID'. To keep type 1 error rates at the

nominal level of 5%, we also included context as a random slope within Signaller ID where

appropriate (Schielzeth & Forstmeier, 2009, Barr et al., 2013). In our sample, as age did not

vary within individuals, it was not included as a random slope. we first assessed whether the

full model explained a significant amount of variation in the response variables by

comparing the full model with a null model containing just the control variables (sex, group,

observation year), random effect, random slope and intercept (Forstmeier & Schielzeth,

2011; Mundry, 2014). To assess the significance of the interaction term we used R function

drop1, with argument 'test' set to 'Chisq'. If an interaction term was not significant, it was

removed. we instead used a reduced model without these interaction terms to allow for

interpretation of the effect of the respective fixed factors. Model comparisons were done

using likelihood ratio tests (LRT; Faraway, 2016) available as R function 'anova' in the

package 'stats'

To examine the influence of our key test predictors on the observed frequency of

unimodal and multimodal combinations, we fitted two models for each response variable.

Signal frequency was based on the number of unimodal or multimodal combinations per

communicative interaction. This included interactions where the frequency of both of these

signal categories was 0, that is, interactions where only unanalysed signal categories were

produced. In the second model, where multimodal combination frequency was the response

variable, the data contained an excess of 0s due to a high number of communicative

interactions where no combinations were produced. To account for this, we constructed a

zero-inflated negative binomial model using the glmmTMB function of the R package

'glmmTMB' (Brooks et al., 2017). This was not a problem for the first model as unimodal

signals were the most recorded signal category. This model was constructed with a Poisson

error structure and log link function using the glmer function from the R package 'lme4'

(Bates et al., 2015). we controlled for variation in observation time across individuals by

including log (total hours observed) as an offset variable in both models.

After establishing overall patterns, additional models were used to look closer at

production of unimodal signals relative to multimodal combinations and variation in the

production of individual signal modalities/multimodal combinations. First, a mixed-effects

binomial logistic regression was used to test whether our test predictors affected the probability

of a multimodal combination being produced instead of a unimodal signal. The binary response

variable was the signal category (0=UM, 1=MM combination). Next, to examine variation in the

production of individual signal modalities as unimodal signals (facial expressions, gestures,

vocalizations), we fitted one model with a Poisson error structure and log link function for each

of the three response variables. The intention was to construct an equivalent model to examine

the production of different multimodal combinations; however, at this time the data lacked

sufficient variation across individuals and contexts to conduct viable inferential statistics (see

Appendix 1.3 Table A1.7). Variation is still described.

Prior to running all our models, we z-transformed the continuous variable 'age' to a mean

of 0 and a standard deviation of 1 to allow for greater interpretability of coefficients given the

inclusion in interaction terms (Schielzeth, 2010). For models constructed with a Poisson error

structure, overdispersion was checked but did not appear to be an issue (maximum dispersion

parameter = 0.85). We assessed the stability of all models using a function written by Roger

Mundry. Estimates obtained from each model based on all data were compared with respective

models with the level of random effects excluded one at a time. This revealed no serious stability

issues across our models. Additional diagnostics for all models are provided in Appendix 1.3.

All models were run in R v. 2.15 (The R Foundation for Statistical Computing, Vienna, Austria,

http://www.r-project.org).

Results

Overview of Signals

Unimodal signals. In total, we observed 1085 unimodal signals across all

individuals and contexts. Unimodal signals were the most produced signal category across

all ages (*N*, individual mean ± SD: infants: 472, 47.20 ± 28.92; juveniles: 424, 60.57 ±

37.24; early adolescents: 189, 17.19 ± 6.95). Across all age groups, facial expressions were

the most common unimodal signal recorded (501, 18.00 ± 18.65) followed by gestures (474,

16.89 ± 12.17) and vocalizations (110, 3.86 ± 4.36). Across contexts, the highest number of

unimodal signals were produced in the play context in all age categories (Table 2.5).

Table 2.5. Individual mean number (±SD) of unimodal signal modalities produced per

communicative interaction (*N* = 807) across age categories and contexts

Age Category (*N*)	Modality	Context			
		Feeding	Play	Aggression	Other
	F	0.13 (0.44)	22.40 (15.45)	0 (0)	0.18 (0.32)
Infant (10)	G	6.10 (4.63)	8.60 (6.65)	0.60 (1.07)	4.40 (5.23)
	V	0.80 (1.23)	4.70 (5.23)	0 (0)	1 (2)

	F	1.14 (1.07)	30.29 (24.62)	0.86 (1.57)	0.14 (0.38)
Juvenile (7)	G	10 (7.46)	6.14 (4.30)	0.71 (1.11)	7.57 (4.54)
	V	0.14 (0.38)	2.29 (4.54)	0 (0)	0 (0)
Early	F	0 (0)	4.18 (4.67)	0.09 (0.30)	0.27 (0.47)
adolescent	G	2.91 (4.09)	1.36 (1.43)	2.27 (3.07)	3 (2.49)
(11)	V	0.09 (0.30)	1 (2.49)	1 (1.26)	0.27 (0.65)

F=facial expression; G=gesture; V=vocalization.

Multimodal signal combinations. We observed a total of 332 multimodal combinations, making them rare relative to unimodal signals (see Appendix 1.3 Fig. A1.1). However, multimodal combinations occurred in 27 of 28 individuals across all age categories (N, individual mean ± SD: infants: 100, 10 ± 9.23; juveniles: 117, 16.71 ± 9.54; early adolescents: 115, 10.45 ± 7.92). we recorded a total of 110 different combinations which included up to six different facial expression types, six different vocalization types and 33 different gesture types (see Table 4 for detailed multimodal combination repertoire). Facial– gestural signals were the most recorded combination (117, 4.18 ± 4.97) and facial– gestural– vocal signals the least (23, 0.82 ± 1.10). Infants and juveniles produced the highest number of multimodal combinations in the play context, but early adolescents produced the most in aggressive contexts (Table 2.6).

Table 2.6. Individual mean number (±SD) of multimodal combinations produced per communicative interaction ($N = 807$) across age categories and contexts

Age category (N)	MM combination	Context			
		Feeding	Play	Aggression	Other
Infant (10)	FV	0.10 (0.30)	4.20 (5.13)	0 (0)	0 (0)
	FG	0.10 (0.30)	3.40 (4.10)	0 (0)	0.50 (1.50)
	GV	0.90 (1.14)	0 (0)	0.10 (0.30)	0.30 (0.90)
	FGV	0.10 (0.30)	0 (0)	0.30 (0.90)	0 (0)
Juvenile (7)	FV	0.71 (1.16)	2.43 (3.33)	0.71 (1.16)	0 (0)
	FG	0.43 (0.73)	7.14 (4.85)	0.43 (0.50)	0.57 (0.73)
	GV	1.86 (2.70)	0.14 (0.35)	0.43 (1.05)	0.86 (1.13)
	FGV	0.43 (0.73)	0.13 (0.35)	0.14 (0.35)	0.23 (0.70)
Early adolescent (11)	FV	0.18 (0.39)	1.63 (3.08)	0.91 (1.16)	0.46 (0.89)
	FG	0.09 (0.29)	0.64 (0.88)	0.91 (0.16)	0.36 (0.88)
	GV	1.00 (2.00)	0.18 (0.59)	2.64 (4.72)	0.36 (0.48)
	FGV	0 (0)	0.18 (0.58)	0.82 (1.47)	0.09 (0.29)

F=facial expression; G=gesture; V=vocalization.

Age and context-related variation in unimodal and multimodal combination production

We ran two models to test the effects of our key predictors on the frequency of unimodal and multimodal combination production. Each model contained 807 data points

corresponding to the total number of communicative interactions observed across 28

individuals. In neither model was the interaction term between age and context significant

and removed from respective models. Overall, the reduced models explained a significant

amount of variation in the frequency of unimodal and multimodal combination production

(LRT comparing the reduced and null model for unimodal signal frequency: UM: χ^2_7

=41.31, $P<0.001$; MM combination: χ^2_7=28.81, $P<0.01$).

In terms of signaller age, we found no significant effect on the frequency of

unimodal signal production (estimate $\pm$ SE=-0.07 $\pm$ 0.05, χ^2_{12}=0.11, P=0.29); age, however,

had a significantly positive effect on the frequency of multimodal combination production

(0.21 $\pm$ 0.09; χ^2_{12}=5.41, P=0.020; Fig. 2.1). While we did not find a significant interaction

between age and context of signal production, overall unimodal signals were produced

significantly more frequently in the play context (0.65 $\pm$ 0.13; χ^2_{12}=18.49, $P<0.001$) and

significantly less in the aggression context (-0.56 $\pm$ 0.16; χ^2_{12}=11.71, $P<0.001$) than the

reference category (Fig. 2.2a). Multimodal combinations were also produced significantly

more in the play context than the reference category (0.70 $\pm$ 0.19; χ^2_{12}=14.42, $P<0.001$) but,

in contrast to unimodal signals, were produced at significantly higher frequencies in the

aggressive context (0.60 $\pm$ 0.21; χ^2_{12}=7.45, P=0.006; Fig. 2.2b). Lastly, we found a

significant effect of our control effects of group and observation year in both models.

Effects of all control variables and nonsignificant key predictors are provided in Appendix

1.3 Table A1.8.

Figure 2.1. Effect of individual age (years) on the frequency of multimodal (MM) combination production in immature chimpanzees ($N = 28$). The mean number of MM combinations produced by each focal individual is shown. Area of the dots reflects the variation in sample size for each individual for each year of age. The solid line and dashed lines represent the fitted GLMM and 95% confidence intervals, respectively.

Figure 2.2. Effect of communicative context on the frequency of (a) unimodal (UM) signals and (b) multimodal (MM) combination production in immature chimpanzees ($N = 28$). Indicated are signal cases (dots), median (horizontal lines), quartiles (boxes), percentiles (2.5 and 97.5%, vertical lines) and distribution of data (shaded areas). Asterisks represent significance (* = <0.05; ** = <0.01; ***= <0.001).

Production of multimodal combinations relative to unimodal signals

I constructed a model to test whether the proportional production of multimodal combinations relative to unimodal signals was affected by our key test predictors. The response variable was binary with one row per individual signal (0=UM signal; 1=MM combination) produced. The model comprised 1417 data points representing all instances of unimodal signals and multimodal combinations produced acoss communicative interactions. Here again, we found no significant interaction between age and context, so the interaction term was removed from the model. Overall, the reduced model explained significantly more variation in the response than the null model (χ^2_7=29.29, *P*<0.001).

Signaller age had a significant positive effect on proportional production of multimodal combinations compared to unimodal signals (0.22 ± 0.20; χ^2_{12}=6.00, *P*=0.014; Fig. 2.3). Again, while we found no significant interaction between age and context (but see Appendix 1.3, Fig A1.2 for effect of age on MM combination production within each context), the proportion of multimodal combinations produced relative to unimodal signals was significantly higher in the aggression context than the reference category (1.58 ± 0.34; χ^2_{12}=13.58, *P*<0.001; Fig. 2.4). Effects of nonsignificant key predictors and control variables are provided in Appendix 1.3 Table A1.9.
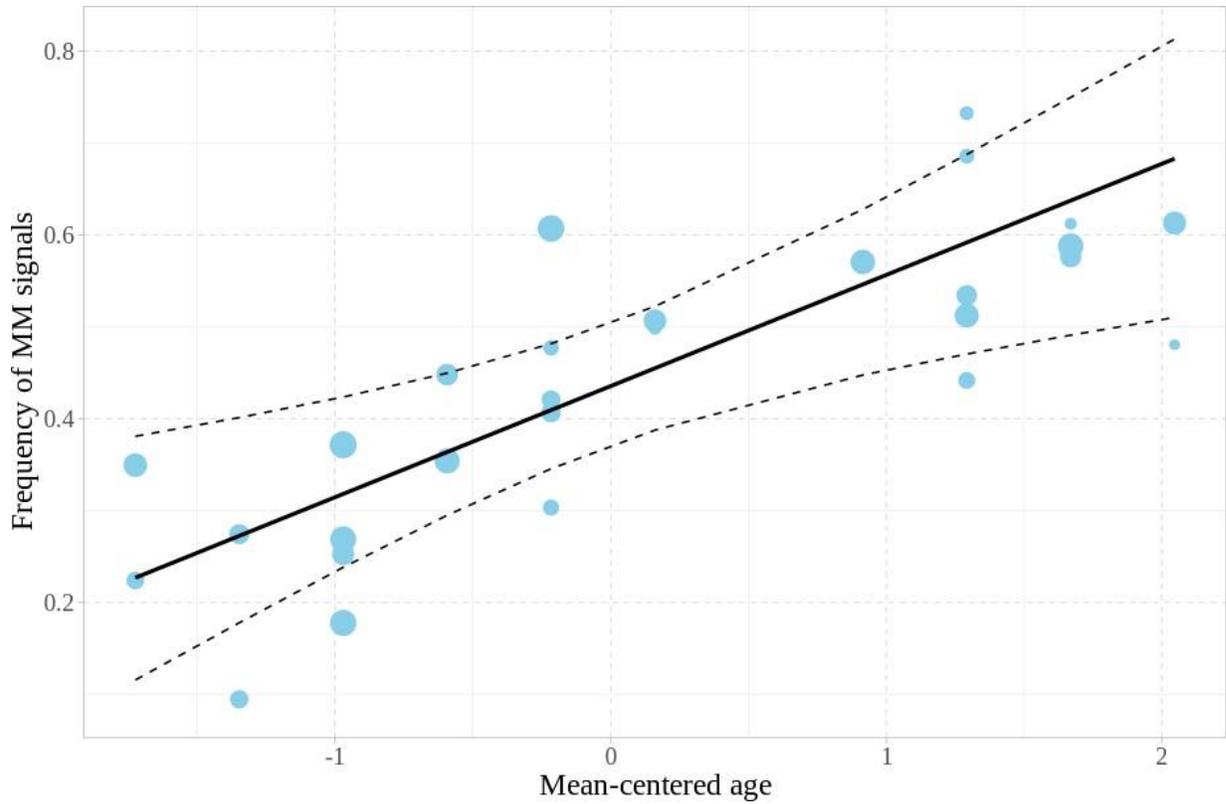
Figure 2.3. Effect of individual age (in years) on the proportion of multimodal (MM) combinations produced by immature chimpanzees ($N = 28$). The proportional production of MM combinations in relation to unimodal signals is shown for each focal individual. Area of the dots reflects the variation in sample size for each individual for each year of age. The solid line and dashed lines represent the fitted GLMM and 95% confidence intervals, respectively.
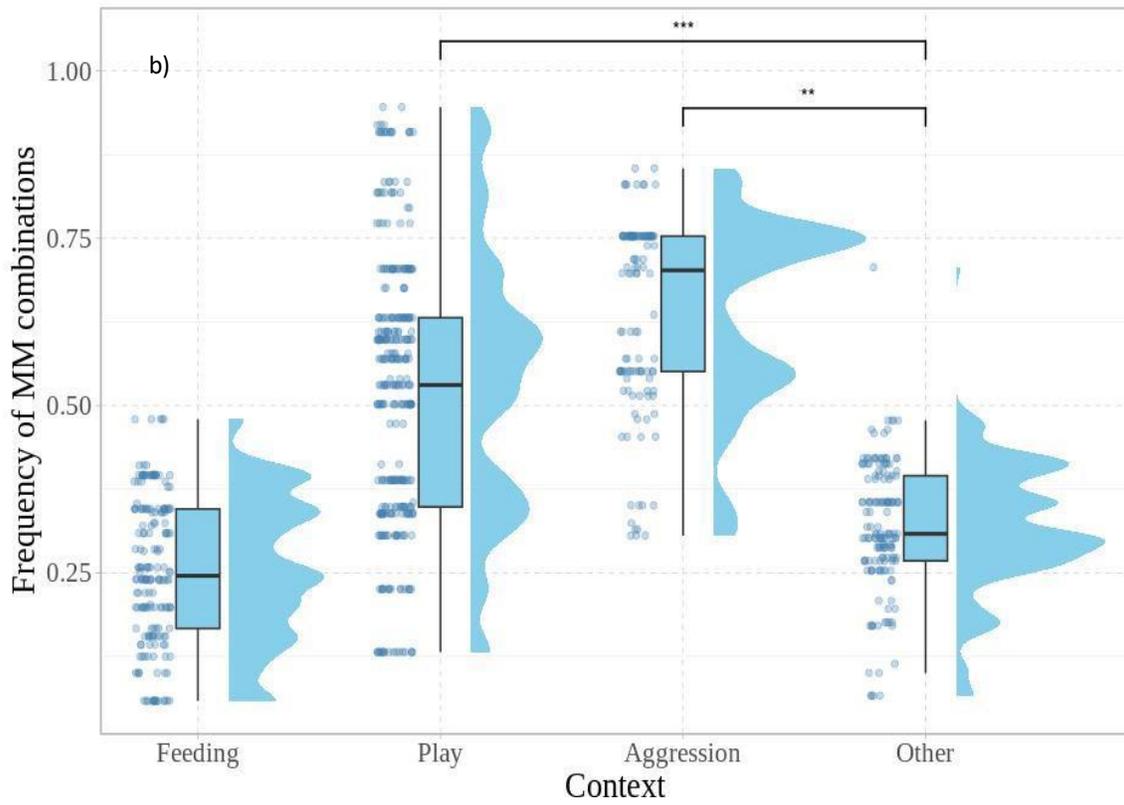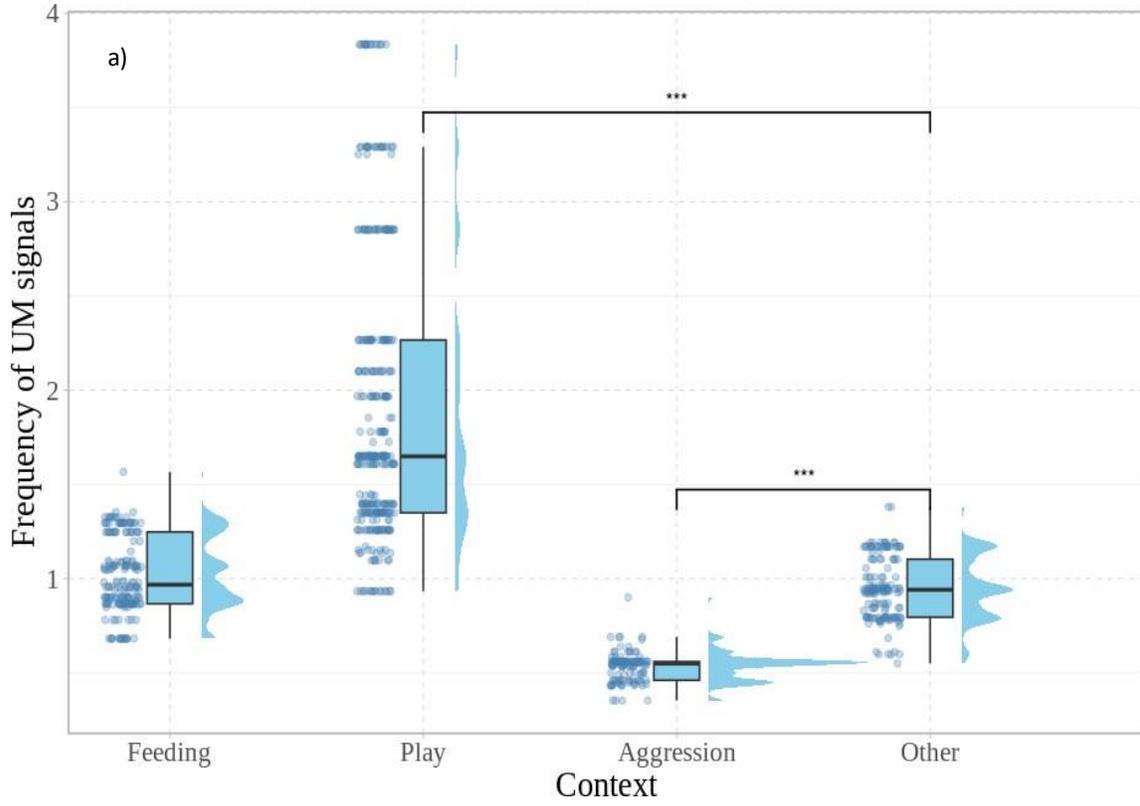
Figure 2.4. Effect of communicative context on the proportional production of multimodal (MM) combinations relative to unimodal signals in immature chimpanzees ($N = 28$). Indicated are signal cases (dots), median (horizontal lines), quartiles (boxes), percentiles (2.5 and 97.5%, vertical lines) and distribution of data (shaded areas). Asterisks represent significance (* = <0.05; ** = <0.01; ***=<0.001).

Variation in the frequency of different unimodal signals

We explored whether our key test predictors impacted the frequency (i.e., number observed per interaction) of unimodal facial, gestural and vocal signals produced by the sample immature chimpanzees. we intended to run three models with the frequency of each modality as a response variable. However, we did not observe sufficient unimodal vocalizations to run inferential statistics on factors affecting the frequency of these signals.

The two remaining models contained 583 data points across all 28 individuals. In neither model was the interaction term between age and context significant. Overall, the reduced models explained significantly more variation in the frequency of unimodal facial expression and gesture production than the null models (LRT comparing the reduced and null model for signal frequency: facial expressions: $\chi^2_7=460.0$, $P<0.001$; gestures: $\chi^2_7=74.187$, $P<0.001$).

In terms of signal modality, age (year) did not have a significant effect on the frequency of unimodal facial (-0.18 ± 0.09; $\chi^2_{12}=3.33$, $P=0.07$) or gestural (0.07 ± 0.06; $\chi^2_{12}=1.42$, $P=0.23$) signals individually. However, there was a significant effect of context on unimodal facial and gestural signal production. Unimodal facial expressions were produced significantly more frequently in the play context than the reference category (2.57 ± 0.27; $\chi^2_{12}=194.981$, $P<0.001$; Fig. 2.5a) in our sample immature chimpanzees. In contrast unimodal gestures were produced significantly less frequently in the play context (-0.73 ± 0.17; $\chi^2_{12}=32.37$, $P<0.001$) and in the aggression context (-0.39 ± 0.20; $\chi^2_{12}=0.049$, $P=0.010$; Fig. 2.5b). we found a significant effect of the control factor of group in the first model. Effects of all control variables and nonsignificant key predictors are provided in Appendix 1.3 Table A1.10.
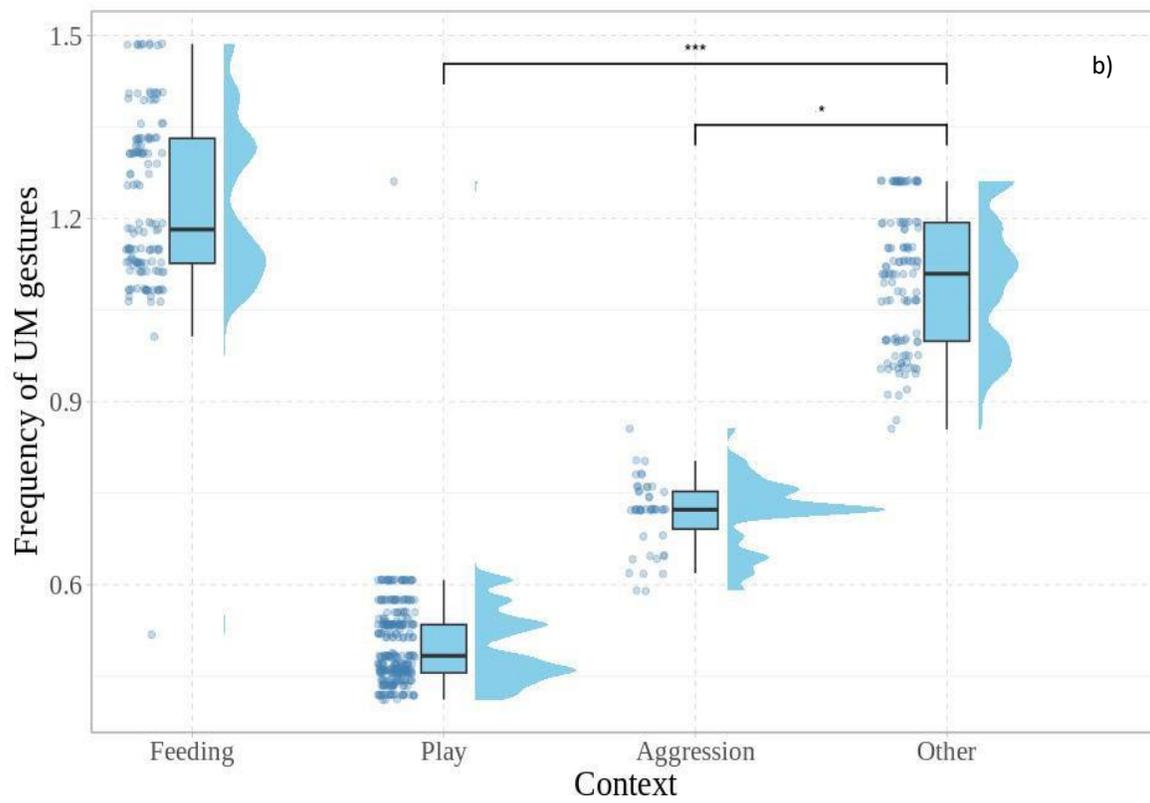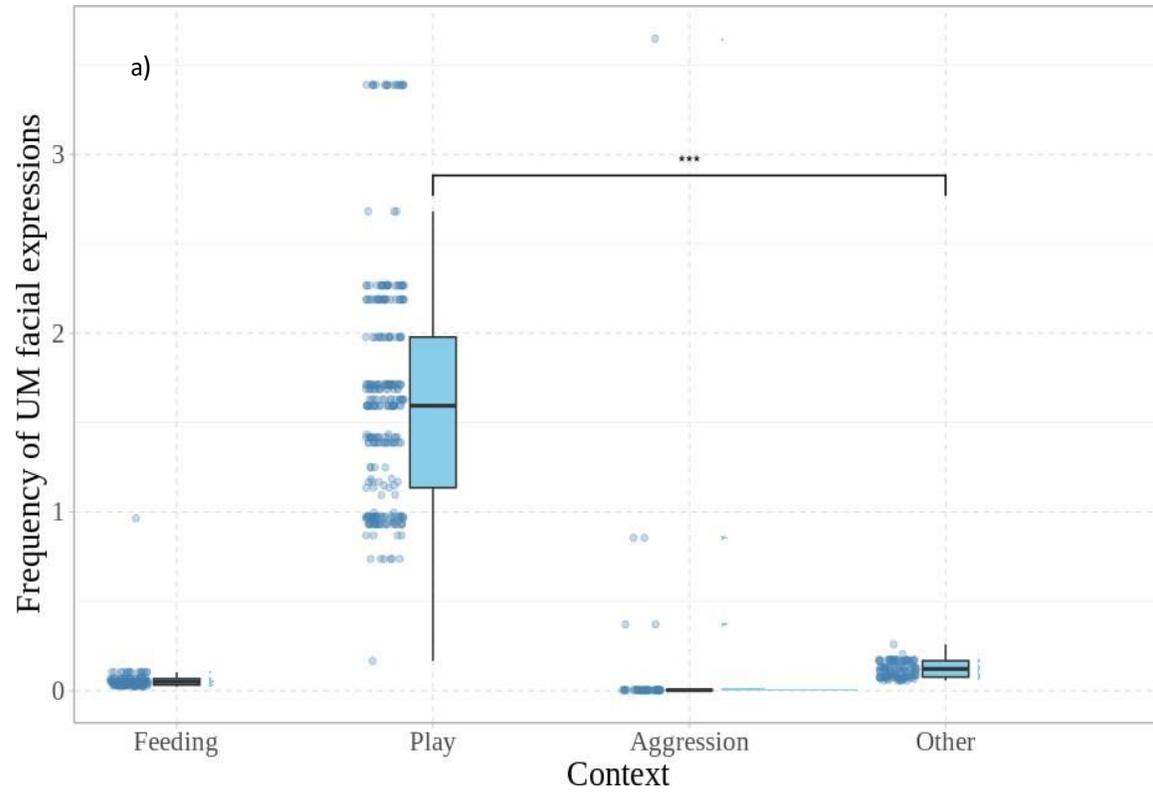
79

Figure 2.5. Effect of communicative context on the frequency of (a) facial and (b) gestural unimodal (UM) signal production in immature chimpanzees ($N =28$). Box plots show the mean number per context (open circles), median (horizontal lines), quartiles (boxes), percentiles (2.5 and 97.5%, vertical lines) and outliers (dots). Asterisks represent significance (* = <0.05; ** = <0.01; *** = <0.001).

Variation in frequency of different multimodal combinations

Lastly, we intended to test whether variation in the frequency of different combinations of multimodal combinations were affected by our key test predictors. However, the frequency of different multimodal combinations lacked sufficient variation across individuals and contexts to conduct viable inferential statistics (e.g., too many individuals had counts of 0 for different combination types and/or 0 multimodal combinations within different contexts). Nevertheless, variation in the frequencies of different multimodal combination types across ages was still notable. Figure 2.6 shows variation across age categories, as variation in the frequencies of combinations was low across some age-years. A full repertoire of multimodal combinations observed across age groups and contexts is provided in Table 2.4.
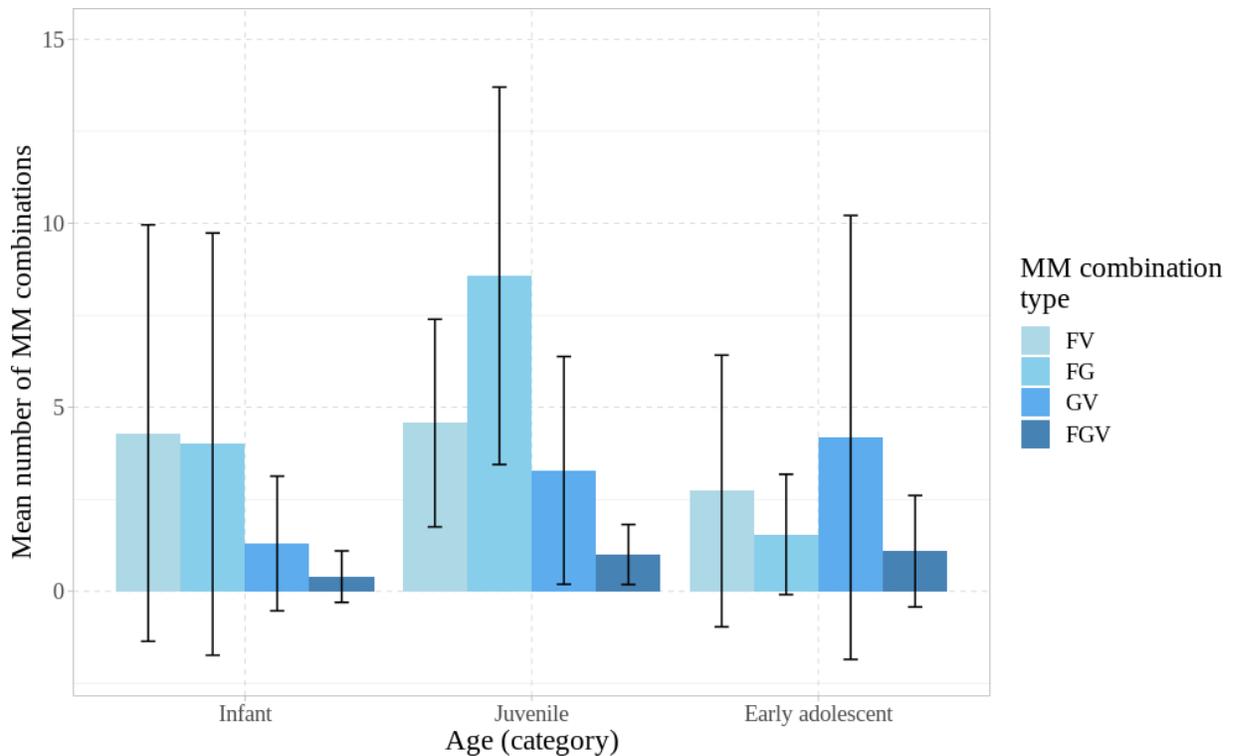
Figure 2.6. Variation in mean number of multimodal (MM) combinations produced across age categories of immature chimpanzees ($N = 28$). Error bars represent mean $\pm$ 1 SD. F= facial expression; G= gesture; V=vocalization.

Discussion

The current study examined the development of unimodal and (free) multimodal communication in immature chimpanzees. Using a relatively large, cross-sectional sample, we found that the majority of immature chimpanzees in our sample (27/28 individuals) produced multimodal combinations, including infants as young as 1 year of age, suggesting that the ability to flexibly combine signals from different modalities occurs early in chimpanzee ontogeny. Importantly, we found significant effects of age: older individuals produced multimodal combinations at greater frequencies and at higher relative proportions than younger individuals, albeit rarely in comparison to unimodal signals. In addition, multimodal

combinations were produced more frequently in the contexts of social play and aggression than in the reference category and were also produced at higher relative proportions than a unimodal signal in the aggression context. Overall, we found a clear difference in the developmental trajectory of unimodal versus multimodal signalling in chimpanzees. The pattern we report, of increasing multimodal coordination, appears to echo that seen in the development of illocutionary communication in human infants (e.g., Iverson, 2010; Gillespie-Lynch et al., 2014), but thus far not systematically examined in nonhuman apes. On the other hand, it also highlights the sustained predominance of unimodal signals, which appears to differ to that seen in humans.

To our knowledge, the current study is the first to specifically focus on multimodal communication development in chimpanzees. Nevertheless, our finding that chimpanzees can produce flexible, multimodal combinations as early as 1 year of age is consistent with some previous related work, particularly from studies of gestural development. For example, Bard et al. (2014) described the addition of facial and vocal components to the gestures of captive chimpanzee infants in their first year of life from as early as 18 weeks in grooming initiations. They also described laughter during tickle play from as early as 8 weeks, but it was not clear whether this included an additional facial component (i.e., open mouth face). Additionally, in a study of joint travel initiation between wild chimpanzee mothers and infants aged 9–69 months, Fröhlich et al. (2016) observed bimodal (gesture plus vocalization) combinations produced by infants at 10 months of age.

Although young chimpanzee infants may have the ability to flexibly combine multiple communicative acts simultaneously, our results indicated that they do so at significantly lower frequencies than older individuals. we believe several developmental factors could be

contributing to this pattern. As chimpanzee infants get older, they become increasingly

spatially independent from their mothers and begin to interact socially with the wider group

(van Lawick-Goodall, 1968). Fröhlich et al. (2017) showed that gesture frequency and

repertoire size of wild chimpanzee infants increased with higher interaction rates with

nonmaternal conspecifics and the number of previous interaction partners, thus highlighting

the importance of interactional experience in communicative development. For our results, it

is possible that the higher frequencies of multimodal combinations produced by older

individuals could (1) reflect an increase in production opportunity, due to increased spatial

independence and interactions with other individuals and (2) be a consequence of a larger

communicative 'data set', that is, a growing signal repertoire from which an individual can

draw a wider range of re-combinable signals. Indeed, while we did not analyse the

distributions of different combination types here, descriptively we found that early

adolescents in our sample produced the largest number of different multimodal combinations

(see Table 2.6). In turn, this larger data set may become increasingly important as older

individuals begin to navigate a more complex social landscape and become increasingly

exposed to more mature interactional social contexts.

Although our results did not reveal the expected interaction between age and context,

we did find that the frequency and relative production of multimodal combinations was

highest in aggressive contexts. Aggressive interactions present obvious personal risks,

including risk of physical injury as well as stress and instability arising from damage to

social relationships. Therefore, in aggressive contexts it may be more important to ensure

messages are communicated clearly and not misunderstood. Previously, ape multimodal

signals have been proposed to function to disambiguate communicative messages (Pollick &

de Waal, 2007; Genty et al., 2014; 2015b, Wilke et al., 2017, Genty, 2019; Oña et al., 2019),

as has been suggested for humans (Partan & Marler, 1999; Vigliocco et al., 2014).

Therefore, the higher rates of multimodal combinations that we observed in the aggressive

context may represent a possible function to disambiguate meaning. we predicted that this

effect would be stronger in older individuals, due to increased behavioural independence

and exposure to competition-induced aggression (van Lawick-Goodall, 1968; Pusey, 1990).

While this age specific prediction was not met, we nevertheless recorded the highest number

of multimodal combinations for early adolescents occurring in the context of aggression

and, further, the number of different combinations produced in aggressive contexts was

substantially higher for early adolescents than younger individuals (see Table 6). For

example, we observed infants performing four different multimodal combinations in the

context of aggression, whereas in adolescents we observed 31 different combinations. This

could have been because older individuals have not only accumulated a larger repertoire of

re-combinable signals via social experiences but also require this wider communicative

range to navigate more complex interactions. At present, this is only speculation. To test

these hypotheses more fine-grained assessments of interactive contexts, considering factors

such as relative age, sex and rank between signaller and recipient, as well as the narrower

context of the aggressive interaction (e.g., access to food/objects, response to threat

displays) are required. Furthermore, it is essential to recognize that the relationships among

these variables are multifaceted, and our models may not have captured all the nuances of

these interactions. We also acknowledge that the absence of significant findings does not

necessarily imply the absence of a genuine effect. The results of this study should be

interpreted in the context of the limitations imposed by our chosen modelling strategies and

data constraints. We remain open to further investigations and potential refinements of the

research design to better understand the interplay between age and context in ontogenetic

patterns of multimodal communication. In turn, this can provide exciting new insight into the degree of functional specificity in different multimodal combinations.

Lastly, we also found that both unimodal and multimodal combinations were produced more frequently in the context of social play. This is not surprising given that play is highly interactive, with individuals producing a range of signals including many open mouth (a.k.a. play) faces, laughter vocalizations and gestures often simultaneously, explaining the high frequencies of multimodal signals in this context. Here, facial expressions were the most frequent unimodal signal type produced in play, again due to the high number of playful open mouth faces. While unimodal gestures were observed significantly less in this context, facial-gesture combinations were observed most often after facial vocal combinations (i.e., open mouth face + laughter) (see Appendix Table A1.7) The open mouth face is an important social regulator in play interactions (Waller & Dunbar, 2005) and is important to make the intention of rougher gestures like hitting or grabbing clear as playful rather than aggressive. Therefore, the function of multimodal combinations within the play context may overlap to some extent with their role in aggressive interactions: to disambiguate the signaller's intended message. Further research into multimodality and its function during play is needed to investigate this.

It is important to acknowledge the influence of the research setting in the interpretation of any study on communication. In an analysis of multimodal communication in wild chimpanzees, Wilke et al. (2017) presented production rates for signals within contexts that occurred frequently enough to be examined further, which included rest, travel and grooming. However, in our study, the majority of signals occurred in the contexts of feeding, play and aggression. This difference in observed contexts of communication could

be influenced by the semiwild context of the sanctuary environment in which we conducted our research. In our study, although the chimpanzees lived in large, forested enclosures, they were only available for observation near the enclosure fence lines; moreover, observations were often associated with periods of artificial provisioning. Food provisioning is likely to increase the frequency of signals related to interspecific food begging as well as those associated with mitigating increased social tension and response to aggression. Play and grooming may also have been more common, given that both can be a form of tension regulation during pre-feeding periods (Palagi et al., 2004). An additional consideration is the potential difference in observation opportunity across each field period. In 2021, there were some added restrictions to researcher movements around the enclosure fence line to reduce risk of transmission and protect chimpanzees from possible Covid-19 exposure. This could have contributed to variation in observation conditions which may thus potentially explain the (control) effects of observation year and group variation in our models. In sum, as the majority of our recorded communicative interactions took place at a period of potentially elevated social tension, our observations may not necessarily reflect patterns of behaviour in wild populations or across other research settings. Although not a factor under investigation in our study, we did also find substantial group level variation in patterns of multimodal production in our samples. Previous studies have also revealed striking levels of group level variation in social tendencies in the chimpanzee groups under investigation here (e.g., van Leeuwen et al., 2018; de Troy et al., 2021). Further investigations of multimodal signal combinations that also take population and context-based variation at an interaction level into account is vital to elucidate communicative patterns and function of multimodality.

Conclusions

Through simultaneously considering facial expressions, gestures and vocalizations in the

analysis of communication behaviours in a sample of immature chimpanzees, our study has provided hitherto undocumented findings of how age and behavioural context affect the production of multimodal combinations at different stages of chimpanzee development. we showed that older individuals use multimodal combination signals at significantly higher frequencies than younger individuals, a pattern that echoes that of illocutionary communication development in humans. In contrast, unlike humans, unimodal signalling remained the dominant form of communication in chimpanzees irrespective of age. These findings highlight the importance of adopting a multimodal approach to primate communication and that by focusing on unimodal signals in isolation, such conclusive developmental patterns can be missed. Moreover, this study provides evidence that behavioural context influences communication behaviour during the immature period, with multimodal combinations potentially acting to add clarity to communicative exchanges where the cost of ambiguity may be higher. Continued investigations that include more fine- grained analysis of interactional context will help to provide critical insight into the functionality of multimodal communication at various stages of ontogeny. Furthermore, attention should be primarily focused on multimodal communication development within and across wild populations to understand the role of the socioecological environment on signal use across time, and thus the selective pressures that may have encouraged multimodality within the hominid lineage ultimately culminating in human language.

Chapter 3 - Maternal interaction behaviour and responsiveness in early infancy

Abstract

Human caregivers systematically adjust their communicative behaviour to compensate for infant developmental stage and corresponding communication skill. Appropriate adjustments in infant-directed communication and response behaviour facilitate infant communicative competency, including language acquisition. Given the widely recognised role that caregivers play in in shaping human communication development, it is therefore surprising that in comparison, we know very little about the role caregivers play in one of our closest living relatives, the chimpanzee (*Pan troglodytes*). The first two years of a chimpanzee's life are characterised by near constant physical closeness to their mothers, with chimpanzee mothers representing the first and primary interaction partner for their infants throughout early development. It is therefore likely that mother-offspring interactional experience does influence chimpanzee communication development. In the present study, my goal was to address this by examining chimpanzee maternal interactional strategies in early infancy (0-2.5 years) including the degree to which chimpanzee mothers respond to the communicative attempts of their infants. Using a sample of 21 semi-wild mother-infant dyads housed at Chimfunshi Wildlife Orphanage, Zambia, I investigated (1) whether mother's interactive behaviour (i.e., physical actions vs communicative signals (gestures, vocalizations, facial expressions)) varied as a function of infant age and (2) whether maternal responsiveness is predicted by infant age and communicative behaviour. Overall, the results show that mothers of older infants used significantly more communicative signals towards their infants than physical actions. In addition, while maternal responsiveness towards their infants decreased with infant age,

chimpanzee mothers responded significantly more to infant signals than to infant physical actions. These findings highlight that modifications in maternal communication behaviour do occur in infant ontogeny and could act to support infant communicative development.

Introduction

Like most primate species, the first and primary social relationship in a young chimpanzee's (*Pan troglodytes*) life is with their mother and the first two years are typically spent in near constant maternal physical contact (Van Lawick-Goodall, 1967). As with other apes, including humans, chimpanzee early life history is characterized by a protracted period of dependence on their primary caregiver (i.e., mother), during which time they learn various skills and knowledge essential to subsequent survival (Londsdorf, 2006, 2012). In addition to the acquisition of technical skills (e.g., nest-building and tool use), maternal interactions are also important for the development of the social skills necessary for navigating the complex, multilevel societies in which they grow up. Even post-weaning, chimpanzee mothers remain an important social partner (Pusey 1983, 1990; Watts & Pusey 2002; Stanton et al., 2017). The detrimental effects of maternal loss on social behaviour and integration have been documented to begin early and persist into adulthood (Botero et al., 2013; van Leuween et al., 2014; Kalcher-Sommersguter et al., 2015).

While well-developed communication skills are imperative for the successful formation and maintenance of primate social relationships (Altman, 1962), research focused on the role of maternal interactions on this aspect of young ape development remains limited. This is surprising given the importance assigned to these early exchanges for other aspects of development and its demonstrated importance in humans. For example,

the critical role of early-interactive contexts in building a communicative foundation for child language development has been studied extensively both in terms of how adults communicate with and respond to the interactional bids of their infants. Given the traditional view of language as speech (Hockett, 1960; Lieberman et al.,1975; Hauser et al., 2002), infant-directed communication has primarily focused on a specialized speech register reserved for infants and toddlers known as infant-directed speech (or IDS, vocal motherese). Across cultures, IDS, and caregiver 'responsiveness' i.e., prompt, contingent and appropriate responses to infant communicative attempts (Ainsworth et al., 1978; Bornstein & Tamis-LeMonda, 1989), have been shown to promote language learning through multiple mechanisms (see for reviews: Tamis-LeMonda et al., 2014; Golinkoff et al., 2015). Furthermore, human caregivers shift the sophistication of their interactional behaviours to accommodate and encourage the developing skills of their infants (e.g., Bornstein et al., 2008; Karasik et al., 2014; Tamis-LeMonda et al., 2013), a process referred to as 'scaffolding'.

In contrast to the wealth of research in humans, we still know comparatively little regarding early interaction behaviour within chimpanzee mother-infant dyads, nor whether chimpanzee mothers also demonstrate similar, albeit less ostensive shifts in their interactional strategies across this important developmental period. In some early descriptions of mother-infant interactions, van Lawick-Goodall (1967, 1968) highlighted the importance of 'manipulative' tactile movements by mothers during the first few months, for example, "pushing" or "pulling" to physically reposition the infant on her body during locomotion or breastfeeding. More recently, in the context of joint travel initiation, Fröhlich et al. (2016) presented evidence that wild chimpanzee mothers increasingly replaced these action-based behaviours with more (visual) signalling to

initiate carries with infants. In this study, infants ranged in age from 9-69 months old, and the reported longitudinal increase in visual signalling was attributed to the natural increase in physical proximity between infants and their mothers during this period.

While chimpanzee offspring remain behaviourally dependent on their mothers until around 8 years of age, during later infancy and the juvenile period, they become increasingly spatially independent, spending longer periods at greater distances from their mother (van Lawick-Goodall, 1968; Behringer et al., 2014). Within this context, it therefore seems evident that pattern of increased visual signalling observed by Fröhlich et al (2016) would align with infants becoming increasingly out of reach to make mechanically effective physical actions a viable option to initiate carrying. However, a notable empirical gap remains regarding maternal communication *prior* to this period of growing physical independence. In seminal work on wild chimpanzee mother-infant interactions, Plooij (1978, 1979) described that similar to human children, infant chimpanzees also undergo a developmental shift at 9-12 months from physical, perlocutionary acts (acts with understood meaning) to more illocutionary communication (acts with intended meaning). Examining variation in maternal interaction behaviour during this period could shed further insight into their understanding of their infants' developmental stage. That is, if increased physical- proximity is removed as an explanation for a shift from mothers physical to signalling behaviour, it could reflect that chimpanzee mothers may instead adjust their infant-directed behaviour in accordance with their infants' abilities and skill levels, similar to human parents.

Moreover, the period of early infancy and the associated close physical proximity between infants and their mothers also provides an interesting period to study changes in

maternal response behaviour. Although caregiver responsiveness is credited with having such an important role in supporting early communicative development in humans, we know very little about its prevalence in our closest living relatives or whether ape mothers also modify responsiveness to infant developmental stage. That said, in a study of mother-infant gestural communication across the four great ape species, that encompassed the period of early infancy (0-20 months), Schneider and colleagues (2016) reported that chimpanzee mothers responded to around just 50% of all infant gestures, demonstrating an apparent 'ambivalence' to infant gestural communication. However, the authors did not consider other types of infant interactive behaviours mothers may have had the opportunity to respond to or not, nor the potential of additional confounding factors which more recent evidence has suggested could influence her decision to do so. For example, recent studies in the vocal domain have indicated that subtle acoustic features of infant distress vocalizations can affect the likelihood of maternal responses (Dezecahce et al., 2020). More specifically, the authors found that acoustic information contained within distress calls (i.e., whimpers) can convey information about discrete problems experienced by the infant (e.g., threat, separation, pain) as well as their distance to their mothers, which may guide parental decisions e.g., to gaze, approach or pick up their infant. Furthermore, Taylor and colleagues (2022) also demonstrated that infant calls accompanied by directed cues (gaze or face directedness) were predictive of behavioural responses in social partners including mothers.

While again these studies only focused on one behavioural type, they both add support for the finding that chimpanzee mothers appear more selective in their responsiveness but add that response behaviour could be more related to the 'risk level' to the infant and recognition that a signal is directed towards them. Thus, the period of early

infancy and associated physical dependence between infants and mothers could also reveal a time of transition from the mother imputing meaning from infants perlocutionary acts, to increasing opportunity to respond accordingly to growing signalling ability (Plooij, 1978, 1979). While maternal responsiveness in chimpanzees appears lower than in human caregiver-infant interactions in general, an approach which also considers the earliest interactive behaviours of infants in the coding process (i.e., perlocutionary action-based behaviours) could still reveal new insights into the types of behaviours mothers respond to, and therefore her role in earliest communication development. That is, if maternal responsiveness is increasingly dictated by infant ability to produce appropriate and effective signals, this could demonstrate that like humans, mother's response behaviour could play a supportive role in the development of infant communicative competence. Studying maternal interactional behaviour across the period of early infancy from both angles, including the behaviours used to initiate interactions as well as responsiveness to infant interactional bids could allow for more comparability between ape mothers and human primary caregivers and reveal that ape mothers also provide a degree of 'scaffolding' in early infant interactions.

Therefore, in the present study I aimed to examine maternal interaction behaviour during chimpanzee early infancy, and the degree to which chimpanzee mothers respond to the interactive attempts of their infants. Using a within-/between-subjects design, I analysed mother-infant interactions using a sample of 21 dyads with infants aged 0-2.5 years. Focusing on this developmental period allows for the examination of changes in maternal interactive behaviour prior to the major development shift in infant spatial independence and therefore test what has been previously credited as the driver of shifts from physical actions to (visual) signalling (e.g., Fröhlich et al., 2016). Specifically, I

addressed the following research questions: First, does variation in maternal interaction behaviour reflect their infant's developmental stage, in relation to their ability to use communicative signalling? To address this question, I considered the means by which mothers attempt to interact with their infants. I was primarily interested in how they use physical actions (e.g., mechanically effective or locomotory acts) in relation to communicative signals (e.g., gestures, vocalizations, facial expressions). Based upon descriptions of infant developmental shifts from more action-based to increasingly communicative signalling at approximately 9-12 months (Plooij, 1978, 1979), I hypothesized that chimpanzee mothers will adjust their interactive behaviour to the infants age-related skill level. I therefore predict that there will be a corresponding shift in maternal interactional behaviour from proportionally more physical actions to communicative signals with increasing infant age.

Second, I examined the hypothesis that maternal responsiveness is also affected by age-related variation in infant interactive behaviour. To test this, using a definition adapted from human literature (Ainsworth et al., 1978; Bornstein & Tamis-LeMonda, 1989), I examined variation in maternal responsiveness i.e., prompt and contingent changes in mothers' behaviour subsequent to an infant's physical act or communicative signal. With limited evidence of maternal response behaviour in chimpanzee mothers on which to base an expectation, I base our predictions on patterns seen within human infant-caregiver interactions, given our close phylogenetic relationship (e.g., Adamson & Bakeman, 1984; Schmidt, 1996; Tamis-LeMonda et al., 2006). To this end, in line with the hypothesis that chimpanzee mothers will adjust to the developing skills of their infant, I predicted that as infants increase in age, chimpanzee mothers will respond to a higher proportion of communicative signals than physical actions.

In addition, given the evidence that maternal interactional behaviour in chimpanzees can be affected by the risk level to the infant as well as increased physical separation (Dezecache et al., 2020), I predict that mothers will (1) use more communicative signals than physical actions and (2) respond more to infants, in higher risk interactional contexts and when the infant is further away.

Note: Originally, I had hoped to include a longitudinal analysis in this study, examining the effect of maternal interactional behaviour in early infancy on measures of communicative skill later in ontogeny. Unfortunately, I did not have sufficient data to include this at this time but is something I would be keen to test with the collection of additional, longitudinal data. Further details and preliminary tests are provided in Appendix 2.1.

Methods

Study site and subjects

Data was collected at Chimfunshi Wildlife Orphanage Trust (hereafter Chimfunshi), a non- profit sanctuary located in the Copperbelt region of northern Zambia (12°23′S, 29°32′E). A detailed description of the study site can be found in Chapter 2. I observed interactions of 21 mothers with offspring ranging from 0-30 months of age (Mean ± SD = 14.73 ± 8.49) (see Table 3.1). Three mothers had young infants in two study periods, so in total I observed 18 mothers and 21 infants.

mothers and 21 infants.

Table 3.1. Description of 21 mother-infant dyads included study sample

Mother-infant dyad	Infant age (months)	Infant sex	Observation year	Observation time (h)

Ingrid-Ivy	1-2[a]	F	2021	1.0
Dora-Dean	1-4	M	2021	2.17
Trixie-Teresa	2-5	F	2021	2.08
Gonzaga-Gay	5-8	F	2021	2.33
BJ-Ben	6-9	M	2021	2.17
Tess-Telma	9-12	F	2021	2.58
Dolly-Duncan	10-13	M	2021	2
Tilly-Tim	10-13	M	2021	1.92
Daisey-Derek	11-14	M	2021	2.33
Diz-Daphne	11-14	F	2017	2.17
Maxine-Mischek	13-16	M	2021	2.25
Ingrid-Ida	14-18	F	2017	3.33
Mischa-Molly	14-17	F	2021	1.75
Genny-Glora	15-19	F	2017	5
Debbie-Don	17-21	M	2021	2.17
Judy-Jake	23-26	M	2021	2.33
Innocentia-Isaac	24-27	M	2021	2.33
Claire-Charlie	24-27	F	2021	1.17
Tess-Tina	24-28	F	2017	4.67
Kambo-Kenny	24-27	M	2013	2
Claire-Chitalu	27-30	F	2017	2.58

a. Born during the field period

Data collection

Recorded observations were made on 20 chimpanzee mother-infant dyads during two main

periods between May-August 2017 and May-August 2021. I also had access to additional

video recordings of one dyad (Kambo and Kenny) taken in June-September 2013, bringing

the total to 21 dyads. Details of sampling approach are consistent with those described in

Chapter 2. A total of 50.39 hours of focal observations were collected (Mean ± SD = 2.39 ±

0.91 hours per mother-infant dyad) (see Table 3.1 for full details of observation times).

Behavioural coding

A total of 274 high-quality recordings were available for coding using ELAN (version

6.0) open-source video annotation software (https://archive.mpi.nl/tla/elan). All instances

of

interaction behaviours produced within the mother-infant dyad were coded. While coding

interactive behaviour, I differentiate between physical actions and communicative signals

(i.e., gestures, facial expressions, vocalizations) (henceforth signals) and combinations of

these (e.g., physical action + vocalization). I define a physical action as any behaviour

produced in pursuit of an apparent interactive goal via direct physical force through the

manipulation of another's body or the movement of one's own body (adapted from Halina

et al., 2013). Definitions of physical actions include types identified in the current study

and some previous descriptions of actions used by chimpanzee mother-infant dyads (see

Table 3.2 for descriptions of observed physical action types). Apparent interactive goals

coded are provided in Appendix 2.2, Table A2.2.

Definitions of all signals (facial expressions, vocalizations and gestures) are based

on the established ethograms and are in included in Appendix 1.3. A communicative

gesture is defined as a directed, non-locomotor and mechanically ineffective movement of

the head, limbs or body where the signaller shows anticipation of a recipient's reaction via

eye gaze and/or body orientation (e.g., Tomasello et al., 1985; Call & Tomasello, 2007; Hobaiter & Byrne, 2011a; Bard et al., 2014). Therefore, physical actions can be distinguished from communicative gestures given they are mechanically effective and/or can involve locomotion (Halina et al., 2013). Lastly, for each physical action or signal, I also coded contextual risk level (two levels: low; feeding, playing, resting, relaxed travel contexts; high: fearful or aggressive situations involving the mother, infant and/ or surrounding group) and maternal proximity (three levels: body contact: body contact, <1m, >1m).

Table 3.2 Physical action types used by chimpanzee mothers and their infants at Chimfunshi Wildlife Orphange.

Individuals' behaviour observed in	Physical action	Description
Infants	Cling on [a]	Infant clings with one or both hands to fur of mother
	Quick approach [c]	Infant moves hurriedly towards mother
	Nuzzle [b]	Infant rubs head from side to side and up and down against mother's body searching for nipple
Mothers	Hold in position [c]	Mother holds her infant in place with the palm of their hand against during repositioning and/ or locomotion
	Shake back [c]	Mother shakes lower back in an upward movement to reposition infant while they are already clinging

Both	Push [b]	Mother/infant forcefully pushes to physically move the body or body part of the other away from oneself
	Pull [b]	Mother/infant forcefully pulls to physically move the body or body part of the other towards oneself

[a] Fröhlich et al., 2016. [b] Nishida et al., 1999. [c] Identified in this study.

Maternal response is here defined as a contiguous and contingent (i.e., dependent on infants preceding act) observable change in behaviour. For example, if an infant whimpers towards its mother while she is feeding, and she then offers her infant a food item, she is credited with responsiveness; similarly, if an infant clings to the fur of their mother while she is standing/ moving and she then picks them up to carry them, she is also credited with responding. Responses included behavioural responses e.g., shares food, groom, play (including physical actions e.g., pushes and pulls) or signals. No response was recorded if there was no visible change in maternal behaviour after an infant produced a physical action or signal. I also coded 'no response' if passive food sharing occurred i.e., excluding facilitating actions (Boesch & Boesch, 1989). This was because it was too difficult to determine whether this was intentional on the part of the mother. For statistical analyses, I only included cases were a clear "response" or "no response" distinction could be made.

Cases where this could not be determined (e.g., when visibility was obscured) were excluded.

In the few cases where a mother produced a physical action + a signal combination (e.g., a push + grunt) I included this as a 'signal' as I am primarily interested in her use communicative signals in relation to infant age. However, in the few cases where a mother

responded to a physical action + a signal combination produced by an infant (e.g., quick approach + whimper), I did not code the response as I could not determine with confidence which behaviour the mother was responding to.

Inter-coder reliability

To assess the reliability of video coding, a second independent researcher coded 10% of the total recordings of mother-infant dyads (28 of 274 focal recordings, 17 of 21 mother-infant dyads) which were tested using the Cohen's kappa coefficient to ensure inter-observer reliability (Altmann, 1974). The mean kappa value obtained for interaction behaviour (i.e., physical action vs behaviour ($\kappa = 0.75$) indicated very good level of agreement and maternal response (yes/ no) ($\kappa = 0.92$) indicated excellent level of agreement. The cappa value for agreement on contextual risk level (low/ high) ($\kappa = 0.78$) and maternal proximity (bodily contact, <1m, >1m) ($\kappa = 1$) also indicated very good and excellent level of agreement, respectively.

Statistical analyses

To test (1) whether maternal interactive behaviour and (2) maternal responsiveness is predicted by infant age and behaviour, I used Generalized Linear Mixed Models (GLMM) (Baayen, 2008) with a binomial error structure and logit link function. To allow full advantage to be taken of coded information, I used a modified dataset for each model corresponding to the response variable after lines with non-available cases for that response variable were omitted, in addition to non-available cases for each predictor. Total cases for each response variable ranged from 5-27 (model 1) and 5-54 (model 2). Given that infant age varied considerably both between and within dyads, to avoid pseudo-replication and distinguish between age-effects across (cross-sectional) or within (longitudinal) infants (i.e., to not erroneously generalize within-subject effects to between-subject effects or vice

versa) I used the method of within-subject centring (van Pol & Wright, 2009). This method creates two age-related variables to be included into each model: between-infant age (henceforth between-age) which represents the average age of each infant consistent across all data points for respective individual; and within-infant age (henceforth within-age) which represents the difference between the infants actual age at the time of an observation and its average age. In model 2, I also included infant interactive behaviour type (two levels: physical action, signal) as a key test predictor. As I assumed that during ontogeny infants would develop more advanced communicative strategies, I also included an initial interaction term between behaviour type and the two variables representing age. To control for potential confounding effects, as control predictors in both models I included infant sex (two levels: male, female), maternal parity (range: 1-5), contextual risk level (two levels: low, high), maternal proximity (three levels: body contact, 0-1m, >1m) and group number (three levels: 1, 2, 4). As a random effect (intercept) I included dyad identity in both models. Information regarding model checks and implementation procedures is consistent with those described in Chapter 2. Diagnostics for all models are provided in Appendix 2.2.

Results

Overall, I recorded a total of 589 interactive behaviours within 21 mother-infant dyads (dyad mean ± SD: 27.62 ± 15.69) from 274 high-quality video recordings. The coding of this data resulted in a total (number of cases) of 289 physical actions (*N (total number of individuals),* individual mean ± SD: mothers: 188(20), 8.95 ± 6.24; infants: 101(18), 4.59 ± 8.79) and 315 communicative signals (mothers: 135(19), 6.43 ± 4.97; infants: 180(21), 8.57 ± 4.95). Included in these cases were 15 physical actions + signal combinations (mothers: 8(4) infants: 7(5)). The vast majority of signals mothers directed towards their

infants were gestures *(*121(19), 5.76 ± 4.62). The remaining signals included vocalizations

($N = 8(5)$) and multimodal signal combinations ($N = 4(3)$). Regarding maternal

responsiveness, I observed a total (number of cases) of 127 maternal responses to infant

interactive behaviours (individual mean ± SD: 6.05 ± 3.98). This number corresponds to a

response rate of 57.42% to infant interactive behaviours. Proportionately, mothers

responded to 25.74% of infant physical actions and 52.22% of infant communicative

signals. In only 14 cases were infants >1m away from their mother at the time an

interaction behaviour occurred.

Effect of infant age on mothers' interactive behaviour

In the first model, I tested the effect of infant age (within-age and between-age), on the

interactive behaviour of mothers (physical actions and signals). The model contained 315

data points, corresponding to the number of behaviours mothers were observed to direct

toward their infants. In N=8 cases, a physical action and a signal occurred simultaneously

(e.g., push + grunt). In this model, these cases were coded as 'signals' because I am

primarily interested in the use of these behaviours by mothers as a function of infant age.

Overall, the full model explained a significant amount of variation in maternal interactive

behaviour compared to the null model (likelihood ratio tests comparing null and full model

(LRT): $X^2 = 318.05$, d.f. = 2, P=0.022).

In dyads involving older infants, mothers used significantly more signals than

physical actions when interacting with them (between-infant age: estimate ± s.e = 0.514

±0.190, $\chi^2 = 7.134$, d.f. = 1, $P = 0.008$) (Figure 1). In addition, signals were more likely to

be used when there was an increase in mother-infant proximity (2.051 ± 0.295, $\chi^2 = 55.82$,

d.f. = 1, P<0.001) and by mothers who have reared more previous offspring (0.354 ±

0.132, $\chi^2 = 5.392$, d.f. = 1, $P = 0.020$). None of the other effects in the model reached significance (see Appendix 2.2 Table A2.3).
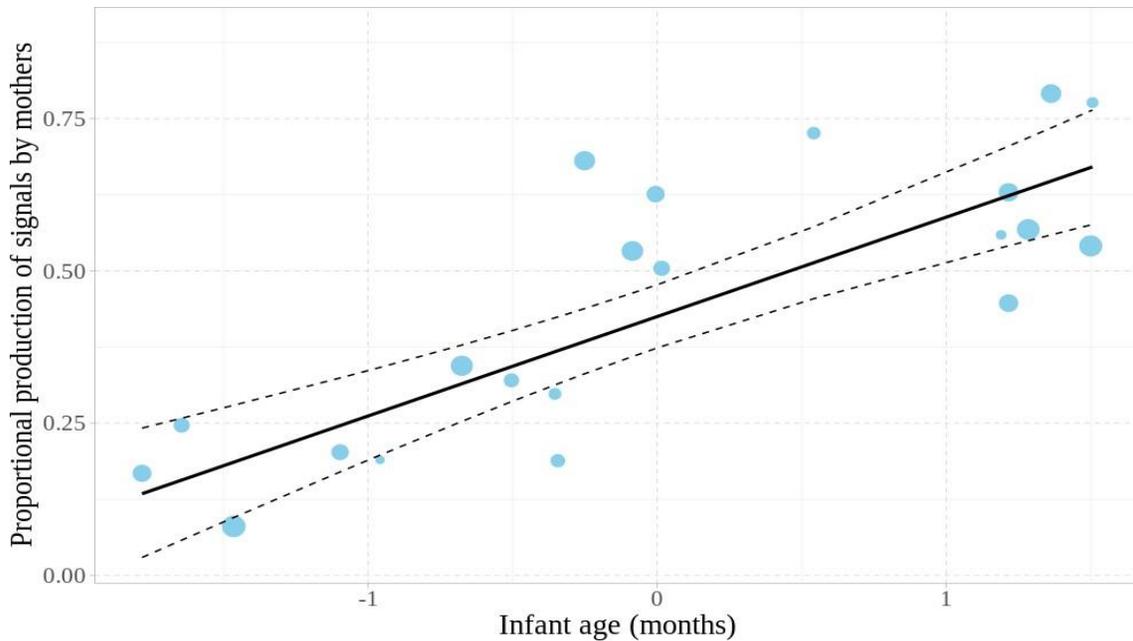
Figure 3.1. Effect of mean-cantered infant age (between) on the proportion of interactive behaviours mothers directed towards their infants that were signals compared to physical actions. The size of the dots reflects the variation in sample size of each individual. The solid line and dashed lines represent the fitted GLMM and 95% confidence intervals GLMM and 95% confidence intervals, respectively.

Effect of infant age and interactional behaviour on maternal responsiveness

In my second model, I tested the effect of infant age (within-age and between-age) and interactional behaviour on maternal responsiveness. The model contained 287 data points, corresponding to the number of infant behaviours mothers had an opportunity to respond to, and a clear distinction could be made in terms of whether she responded or not i.e., excluding unknown cases. In N=7 cases, a physical action and a signal occurred simultaneously (e.g., pull + whimper). Here, these cases were excluded from the model as I was unable to establish

103

which of these components a mother did/ did not respond to. I initially included interaction terms between the age variables (between- and within-age) and infant behaviour type, but these were not significant. A reduced model excluding these interaction terms explained a significant amount of variation in maternal responsiveness compared to the null model (likelihood ratio tests comparing null and full model (LRT) X^2=12.812, d.f. = 3, P=0.005).

In the reduced model, I found a significant effect of within-infant age on maternal responsiveness (-0.395 $\pm$ 0.159, χ^2 = 6.441, d.f. = 1, P=0.011). That is, as infants get older, mothers are less likely to respond to their interactive behaviours (Figure 3.2). However, the type of interactive behaviour employed by infants had a significant effect on the likelihood of eliciting maternal responses. Specifically, infant communicative signals were significantly more likely to elicit a maternal response than physical actions (0.839 $\pm$ 0.345, χ^2 = 5.968, d.f. = 1, *P*=0.015) (Figure 3.3), regardless of infant age. In addition, mothers were more likely to respond to their infants when they were further away (1.588 $\pm$ 0.312, χ^2 = 19.576, d.f. = 1, *P*<0.001) and in more high-risk contexts (1.327 $\pm$ 0.592, χ^2 = 5.691, d.f. = 1, *P*=0.017). None of the other effects in the model reached significance (see Appendix Table A2.4).

Figure 3.2. Effect of mean-centred (within) age on the proportion of maternal responses to infants' interactional behaviours (physical actions or signals). Raw proportions are presented separately for each centred month of age of a given individual; thus, individuals can be represented more than once. The area of the dots corresponds to the sample size per centred month of age. The solid line and dashed lines represent the fitted GLMM and 95% confidence intervals GLMM and 95% confidence intervals, respectively.

Figure 3.3. Proportion of infant interactive behaviour (physical actions vs signals) responded to by mothers. Indicated are individual observations (dots), distribution of data (shaded area), median (horizontal lines), quartiles (boxes), and percentiles (2.5 and 97.5%, vertical lines) across each behaviour type.

Discussion

The current study examined early interactional behaviour including maternal responsiveness in sample of 21 chimpanzee mother-infant dyads living in a semi-wild environment. My primary aim was to investigate whether chimpanzee mothers adjust their interactional bids and response behaviour to engage in increasingly communicative interactions as their infants age. Addressing a dearth of research into this area, I focused on mothers of infants aged 0-2.5 years, a period where infants remain in near-constant physical proximity (van Lawick-Goodall, 1967). This allowed me to examine changes in maternal communication behaviour prior to the developmental increase in infant spatial independence, which has been previously credited as the driver of shifts from physical actions to (visual) signalling (e.g., Fröhlich et al., 2016). Overall, I found that mothers of

older infants use more communicative signals (i.e., gestures, vocalizations, signal combinations), rather than mechanically effective physical actions during their interactions with their offspring, which appears to correspond to the reported developmental trajectory of increasing illocutionary communication in infants (Plooij, 1978). Moreover, the results indicate that maternal responsiveness to all infant interactional bids (i.e., physical actions and signals) decreases with infant age but that regardless of age, mothers respond more to infant signals than physical actions. This study provides novel evidence of developmental adjustments in the interactional strategies of chimpanzee mothers that offers an important point of comparison for studies of human communicative development. Although not directly equivalent to IDS, in a manner that echoes the patterns seen in human caregivers (e.g., Murphy & Messer, 1977; Schaffer et al., 1983; Adamson & Bakeman, 1984; Dimitropoulou et al., 1984; Schmidt, 1996; Tamis-LeMonda et al., 2006), chimpanzee maternal behaviour appears to increase in its communicative complexity in response to infant ontogenetic stage.

Shifts from manipulative, physical actions to signalling behaviour have been described previously from observations of infant-directed communication in chimpanzee mothers (van Lawick-Goodall, 1967, 1968; Fröhlich et al., 2016), a pattern typically explained as a response to concurrent developmental increases in infant spatial independence. Here, I showed that this behavioural adjustment begins earlier in ontogeny, before infants begin to leave the relative physical security of their mothers. I recorded only a few cases of maternal signalling in which the infant was more than 1m away from its mother, but still, mothers of older infants used more signals than physical actions. Interestingly, almost all communicative signals produced by mothers towards their infants were gestures. This is in stark contrast to the literature on infant-directed communication in

humans in which infant-directed speech (IDS) is the primary mode of behaviour studied in the context of early infant-adult interactions. However, while rarer, 'gestural motherese' has been studied in humans (see for review: Özçalışkan & Dimitrova, 2013) where it has been found that gestures are also used by adults when communicating with infants, but that these are most often combined with IDS to reinforce verbal messages (e.g., "look" + point) (Iverson et al., 1999).

The co-speech gesturing included in infant-directed communication is indicative the tight motor-vocal linkages in humans (McNeill, 1992), and the need for human infants to develop this skill for successful language acquisition (Iverson, 2010). Interestingly, the type of gestures that human parents use with IDS start off simpler in form (Bekken, 1989; Iverson et al., 1999) such as deictic (e.g., pointing) and conventional (e.g., shaking the head to mean no) gestures but become increasingly incorporated with complex iconic gestures as the infant develops (Özçalışkan & Goldin-Meadow, 2011) eventuating in mature co-speech gestures. Many of these are distinct from more simple gesture forms and inherently absent in the great apes (see for review: Byrne & Cochet, 2017). Therefore, the shift from action- based to gestural signalling observed here in chimpanzee mothers, resembles the shift from simpler, to increasingly complex (speech-gesture combinations) signals in humans while also highlighting important differences between the species. In humans, (non-sign) gestures mainly support the vocal medium (McNeill et al., 1994; Kendon, 2004), but in chimpanzees the high instances of gestural signalling by mothers could be reflective of their primary role in close-distance communication in great apes (Genty et al., 2014; Hobaiter et al., 2017; Wilke et al., 2017; Fröhlich et al., 2019). In sum, our results indicate that chimpanzee mothers, like humans, use increasingly sophisticated forms of communication with their infants which reflects the communicative medium

which takes precedence in the mature communication systems of each species.

Although my results did not reveal the expected age-related interaction with type of infant interactive behaviour on maternal responsiveness, I did find that overall maternal responsiveness decreased with increasing infant age. This finding contrasts with patterns seen between human infants and caregivers. In fact, while response behaviour varies as a function of human infant developmental stage (e.g., Bornstein et al., 2008; Karasik et al., 2011; Karasik et al., 2014, Tamis LeMonda et al., 2013), on the whole, infant age does not dictate the likelihood of caregiver responding (e.g., Bornstein & Tamis LeMonda, 2008, except in the play context where maternal responsiveness decreased between 14-21 months). However, important species differences in maternal investment and life-history could also help explain this disparity.

In comparison to chimpanzees, human infants are weaned early (see for review: Humphrey, 2010) and while they remain nutritionally dependent on caregivers for many years, human childcare is cooperative: both caregivers and offspring are supplemented with resources and care from other group members (e.g., Hawkes et al., 1998; Reiches et al., 2009; Kramer & Ellison, 2010; Hooper et al., 2015). In contrast, studies from different primate species have highlighted the consequences of an exclusive energetic burden of lactation and parental care on mothers, such as reduced maternal condition, reproductive success, and offspring survival (see for review: Beehner & Lu, 2013). In chimpanzees, most of this energetic demand is concentrated in the first two years of infant life, which encompasses highest relative gains in growth and brain development (e.g., Leigh & Shea 1996; Marzke et al., 1996; Hamada & Udono 2002, Leigh, 2004). Previously, it has been shown that chimpanzee mothers can mitigate the severity of costs incurred to reproductive

fitness during this period via increasing energetic intake (Londsdorf et al., 2012). As chimpanzee infants become increasingly capable of subsidizing their own nutritional requirements (Clark, 1977) from as early as 4-5 months (Rijt-Plooij & Plooij, 1987), mothers need to prioritise energy intake and/ or reservation when associated lactational demand begins to decrease, could contribute to a reduction in attentiveness and by extension responsiveness to infants. This aligns with our finding here that maternal responsiveness was significantly lower in 'lower-risk' contexts, but higher when there was increased proximity between dyad members. That is, as infants get older and more nutritionally independent, mothers could become increasingly selective in responding to her infant unless their perceived well-being was immediately threatened, so to prioritise her own energy intake and reservation.

I also found that chimpanzee mothers responded significantly more to infant signals than physical actions, irrespective of infant age. Here, the most parsimonious explanation seems to be that signals, more so than physical actions, are more easily recognizable as an interactive bid. Physical actions by their nature require function to be inferred through the context in which they are produced. For example, 'pushing' or 'pulling' actions could be attached to many apparent functions or goals across different circumstances such as 'reposition,' 'give me that,' 'stop that,' 'come here' etc. While communicative signals also show varying degrees of functional flexibility (gestures: Tomasello et al., 1994; Liebal et al., 2004; Pollick & de Waal, 2007; Tomasello & Call, 2007; vocalizations: Dezecache et al., 2019, 2021; reviewed in Taylor et al., 2022), in general intended meanings are clearer, for example, a 'mouth stroke' gesture to solicit food sharing, 'screaming' to indicate distress or a 'play face' to clarify playful intentions. Nonetheless, while signals were responded to more, only 52% of signals were responded to by mothers in our sample. This

110

finding is consistent with the only other study of maternal responsiveness in (non-human) apes I am aware of in which Schneider et al. (2016) reported that zoo-housed chimpanzee mothers of infants also aged 0-2.5 years, only responded to 50% of infant gestural requests.

Another possible explanation for low maternal responsiveness to infant signals could also be related to the ontogeny of infant intentional communication. While intentionality in chimpanzee communication is proposed to begin at approximately 9 –12 months (Plooij, 1978), conclusive evidence for this is still missing. It is also likely that intentional signal usage continues to develop outside of my sample age range. For example, in 2018, Fröhlich and colleagues reported that markers related to intentional signalling including audience, checking, sensitivity to recipient's attentional state and persistence to the goal, were sensitive to within- or between-age affects in wild infant chimpanzees aged 9-78 months. It could be that while infant signals maybe more likely to be recognized as communicative, such low maternal responsiveness might reflect young infants underdeveloped ability to communicate effectively. An interesting line of future investigation could therefore be to examine maternal responsiveness to infant signals with or without different markers of intentionality over time, such as visual checking and monitoring of mother's behaviour. The timing and onset of intentional communication in early infancy could be influenced via a such a trial-and-error system of infant communicative attempts meaning that maternal responsiveness could have a scaffolding role in infant communicative development through which infants learn how to communicate effectively.

In this study, I did not directly study the intentionality of infant interactive

behaviour as this was difficult to assess in terms of mechanically effective physical actions, thus I have no measure of whether this changed across age. In terms of communicative signals, gestures are more intentional by definition (Byrne et al., 2017), as they must show evidence of directedness and/ or response anticipation to distinguish them from a non- communicative mechanical or locomotory movement. Facial expressions and vocalizations, however, do not require the same indicators of intentionality to be considered 'communicative' and they are typically assumed to be bound to certain affective states (see for review: Liebal & Oña, 2018). However, there is evidence that chimpanzees are capable of having at least some control over their facial (Davila-Ross et al., 2015) and vocal (e.g., Crockford et al., 2012; Schel et al., 2013; Watson et al., 2015 Dezecache et al., 2021; Leroux et al., 2021) signals. Furthermore, recent evidence has shown that an infant's ability to direct vocalization to their mothers has a positive effect on the likelihood of her responding (Taylor et al., 2022). Separating signals which did or did not show signs of intentional production across modalities should be addressed in future work, to contribute to our understanding of the development of intentional communication in chimpanzees.

In addition, as physical actions here include mechanically effective and locomotory movements, it is difficult to discern whether they are orchestrated intentionally or are intended to be communicative at all. For example, if an infant 'pulls' the arm of their feeding mother towards themselves and steals food from it, while this has a clear interactive goal (i.e., to get a piece of food), it is hard to know whether the infant intends the mother to be an active participant in achieving this outcome or if they are more objectively focused on the outcome itself. However, here too maternal response behaviour could shed light on the transition from these action-based behaviours to increases in

intentional (gestural) signalling in infants. For example, the proposed developmental process of ontogenetic ritualization (Call & Tomasello, 2007) posits that individuals learn gestures in the context of regularly occurring dyadic interactions such that parts of fully functional social behaviours become ritualized. However, here, while I did not look at response behaviour on a context-by- context basis, I found that responses to actions were significantly lower than to signals, which still only occurred around 50% of the time.

For ontogenetic ritualization to occur, recipients must help shape the shift from a physical action to 'ritualized' gesture via their response behaviour, by helping the communicator learn which component of the original action is sufficient to initiate an interaction (Tomasello et al., 1994; 1997). The lack of maternal responses to physical actions here does not conform with this process. At least in the period of early infancy, the developmental shift from actions to more communicative signalling in chimpanzee infants does not appear to be explained by ontogenetic ritualization. Instead, my findings seem more consistent with previous research suggesting that genetic transmission is the major mechanism of gestural acquisition in apes (Genty et al., 2009; Hobaiter & Byrne, 2011b). However, given additional evidence that infant gesture types are used to achieve more specific age-related social goals than mothers (Schneider et al., 2012) and are subsequently shaped by interactions within the wider community (Call & Tomasello 2007; Halina et al. 2013 Fröhlich et al. 2016b), it may also be that the subsequent onset and developmental pathway of (gestural) signal acquisition are reliant on the increasing degree of physical independence from mothers. That is, gestures used in early infancy are more genetically predisposed, but the acquisition of gestures used in a more complex range of social interactions outside the mother-infant dyad is subject to additional learning processes e.g., ontogenetic ritualization. This highlights the importance of considering that different

mechanisms might be responsible for different aspects of communicative development in apes (Liebal et al., 2019) and the inclusion of even the earliest forms of communicative behaviour in infants can aid our ability to disentangle the role of distinct processes operating at different stages of development.

Finally, I need to acknowledge that my largely cross-sectional sample limits my ability to explore within-dyad changes across the whole early-infancy period. While I included a within-age variable in my analyses, this only represented a brief longitudinal period of 3 months. Therefore, the lack of longitudinal effect on maternal interactive behaviour could potentially reflect a lack of change within this short period, but doesn't necessarily exclude the possibility that maternal behaviour is adapted over longer developmental periods. Indeed, the finding that when viewed cross-sectionally, mothers of older infants used more communicative signals than physical interactions could indicate such an adaptation. As such, the results presented here could be built upon by tracking changes in infant and mother communication across the entire period of infant physical dependence; and further, studying how these changes influence subsequent communicative skills e.g., effectiveness and success, could allow more comparability with human literature and the importance of ape mothers as interaction models. Moving forward, longitudinal investigations would allow us to better understand the role of the mother in early interactive contexts and the effect of maternal signalling behaviour and responsiveness on communicative outcomes in chimpanzee infants.

Conclusions

In this analysis of early maternal communication behaviour in chimpanzees, I have shown that mothers also undergo a corresponding shift from physical actions to increased

communicative signalling in a pattern which reflects the trajectory of communication development in infants. Furthermore, evidence that this behavioural shift occurs during the first 2.5 years of infancy, suggests that increases in distance between mothers and infants in later ontogeny is not the only explanation for increases in maternal signalling. Instead, chimpanzee mothers may attune to the changing skills of infants, to become more communicatively intentional given the increasing role that their infant can play as a genuine interaction partner. I also reported that the chimpanzee mothers responded less as their infants got older, but they responded more to infant signals than physical actions. However, in contrast to human caregivers, responsiveness was low overall. The exclusive energetic burden of lactation and parental care could mean that mothers become less attentive to prioritise their own energy reserves as infants begin to become more nutritionally independent. In addition, patterns of maternal responsiveness regarding infant signals could be affected by infant ability to communicate effectively. Future work would benefit of a more fine-grained investigation of the relationship between infant communicative skills with regard to markers of intentional communication and maternal responsiveness. This would help to clarify the potential scaffolding role of maternal response behaviour via encouraging the development of social competence and effective communication in early development.

Chapter 4 - Influence of interaction patterns on multimodal signalling frequency and repertoire size

Abstract

Multimodality is a core feature of human language, which has deep evolutionary roots. Comparative research has revealed that primates, including our closest great ape relatives also produce multimodal signal combinations in their everyday repertoires, indicating that the socio-cognitive prerequisites for multimodality were already present in my last common ancestor. To understand the selective pressures which may have encouraged the evolution of hominid combinatorial signalling, it is necessary to examine the contributing role of socio-ecological environment. However, this aspect has thus far been overlooked. Here, I addressed this by examining the effect of interactional patterns on the development of multimodal communication in immature chimpanzees, based on a cross-sectional sample of 28 semi-wild immature chimpanzees (aged 1 –11 years). Based on my hypothesis of increased multimodality with increased social interaction experience, I investigated whether the quantity of communicative interactions and communicative partners positively predicted the frequency and repertoire size of multimodal signal combinations. I also examined the influence on interactional context on multimodal repertoires, which could inform on the degree of context specificity of multimodal combinations. Contrary to my hypothesis, older individuals with higher social interaction rates produced fewer multimodal combinations per interaction, which may alternatively support a pattern of 'repertoire tuning', previously proposed for chimpanzee gestures. However, in support of my hypothesis, individuals who interacted with a wider range of conspecifics used multimodal combinations more frequently and had larger repertoire sizes, regardless of age. Regarding context, the largest number of distinct combinations occurred during play and aggression, two contexts where it may be

more costly to use ambiguous signals. While unable to test statistically, I also detected high levels of context- specificity, with most multimodal combinations occurring in just one context: indicating that such signals may fulfil specific functions. Overall, results suggest that multimodal development may be positively influenced interactional behaviour, and that large repertoire sizes may play a key role in adding clarity in 'high stakes' interactions. Although multi- modality may enhance communicative complexity, my results also reveal it may serve specific functions.

Introduction

In the study of human communication development, the role of early social experiences in shaping language outcomes is well-established (Hoff, 2006). Primarily, the prolonged period of human infant dependence on primary caregivers (Rosenberg, 2021) means that most early interactive exchanges occur within these relationships which are critical to advance infant understanding of what communication is and how it works (see for review: Renzi et al., 2017). As development proceeds, subsequent socialization with peers then provides further opportunity for a variety of individual and multi-party interactions, including to practice their pragmatic communication skills, which they need to express themselves successfully (Dunn & Shatz, 1989; Ervin-Tripp, 1991; Oshima-Takane et al., 1996; Pellegini et al., 1997).

In contrast with human infants, chimpanzees spend shorter periods relying on their mothers for physical and nutritional security (Humphrey, 2010). While they remain behaviourally dependent on her until adolescence (Pusey, 1990), they begin to leave the relative security of their mother's immediate proximity at around 2 years of age, increasingly integrating with members of the wider social community such as siblings or peers (van

117

Lawick-Goodall, 1986; Pusey, 1990). As such, in comparison with humans, young

chimpanzees have increased opportunity to interact with group members outside of the

mother-infant dyad and subsequently, the opportunity to practice the communicative skills

required to navigate the complex social landscape into which they are to integrate. However,

despite evidence that social experiences strongly influence socio-cognitive development (see

for reviews: Bard & Leavens, 2009; Leavens & Bard, 2011), this has rarely been considered

in terms of its influence on communication development in great apes.

      That said, recent work in the gestural and vocal domain have provided insights

into the role of interactional experience on these aspects of communicative

development in chimpanzees. Concerning gestural signalling, Fröhlich et al. (2017)

expanded on previous descriptive evidence (Plooij, 1978; Maestripieri & Call, 1996) by

showing that with increasing age, higher social interaction rates and more social

partners outside of the mother-infant dyad positively affected different matrices of

gestural signalling including gesture frequency, production of gestural sequences and

repertoire size. More recently, Taylor and colleagues (2022) assessed the impact of

directedness cues and mutual engagement during social interactions on vocal ontogeny

in infant and juvenile individuals and found that like young humans, young

chimpanzees also show evidence of these behaviours in their vocal communication.

Furthermore, aspects of directness in vocal signalling appeared to increase across

ontogeny, reflecting an increasing social competency by producing more effective

vocalizations. While these studies demonstrate the value of considering interactional

experiences in understanding the development of great ape communication systems,

continued work is needed to allow for more direct comparability with human

communication and our understanding of the social pressures shaping its phylogenetic

development.

A historical view of language as speech (Vigliocco et al., 2014) is reflected in an investigatory bias towards the vocal modality in studies of its development (Hoff, 2013) and evolution (Liebal et al., 2014). However, over the last decade there has been growing recognition that human language is an inherently multimodal system wherein speech is routinely and universally paired with additional information contained within eye gaze, facial expressions, gestures and bodily postures. As a consequence, there has been an increase in the incorporation of multimodal perspectives to human communication development, which has tracked an increasing degree of control over the coordination of separate modalities in infants, before the emergence of synchronous gesture–speech combinations (e.g., Esteve-Gibert & Prieto, 2014; Morgenstern, 2014; Murillo et al., 2018). The emergence and usage of multimodal (gesture-speech) combinations has been linked to learning opportunities from social interactions (e.g., Liszkowski et al., 2008; Liszkowski, 2011; Igualada et al., 2015; Morgenstern, 2023) and positively predicts language measures in later development (e.g., Igualada et al., 2015; Murillo et al., 2018; Murillo & Casla, 2020).

With regard to language evolution, comparative research of our closest, great ape relatives has also begun to embrace multimodal approaches to communication, although a dominance of unimodal (i.e., studies of communicative modalities in isolation) studies persists (Liebal et al., 2022). We now know that other apes also routinely use multimodal signal combinations (i.e., temporally overlapping combinations of distinct communicative acts; hereafter interchangeably referred to as multimodal combination) as part of their everyday repertoires (Wilke et al., 2017; Chapter 2) which act to disambiguate

communicative messages (e.g., e.g., Genty et al., 2014, 2015b; Genty, 2019; Hobaiter et al., 2017; Wilke et al., 2017; Oña et al., 2019). However, here too ontogenetic perspectives are crucial yet still lacking. Previously, (Chapter 2), I described patterns of multimodal combination production in a cross-section of semi-wild chimpanzees aged 1-11.5 years. I found that while unimodal signals (i.e., singularly produced gestures, vocalizations, or facial expressions) remain the dominant form of communication regardless of age, multimodal combinations were produced significantly more by older individuals. Moreover, multimodal combinations were produced most often, and at higher probabilities relative to unimodal signals within the context of play and aggression, indicating they are associated with contexts where the cost of ambiguity is heightened e.g., receiving physical aggression. What remains unknown, however, is the degree to which factors related to the interactional environment contribute to this pattern of increasing multimodality. Recent unimodal studies have provided evidence that interactive experience is important for shaping chimpanzee communicative development (Fröhlich et al., 2017; Taylor et al., 2022). However, given the now recognized importance of multimodal approaches in studies of language development and evolution, it is important that this shared feature of our communication systems is considered. Such studies are valuable as developmental processes underlying multimodal communication in apes can inform our understanding of the socio-ecological pressures driving its elaboration into language across evolutionary time.

In this study, I therefore aimed to examine how multimodal communication in immature chimpanzees is associated with matrices related to interactional environment. Specifically, using a cross-sectional sample of 28 immature individuals ranging in age from infancy to early adolescence, I investigated the effect of communicative interaction

rates and number of interaction partners on the use of multimodal signal combinations at different developmental stages, as measured by the frequency of production and repertoire size, as established measures of ape signalling (e.g., de Waal, 1988; Liebal et al., 2004; Call & Tomasello, 2007; Hobaiter & Byrne, 2011a,b) . Given evidence from chimpanzee gestural and vocal development, and the contribution of early interactions to communicative development in humans, I expected that with increasing age the frequency and repertoire size of multimodal signals will be positively related to higher interaction rates and number of communicative partners.

In addition, while multimodal signals in this sample were produced most frequently in the contexts of play and aggression (Chapter 2), here I also considered the effect of interaction context on individual multimodal combination repertoires. Combined with previous evidence from studies of multimodal communication in *Pan* which has demonstrated that signal combinations function to disambiguate communicative messages (e.g., Genty et al., 2014; Hobaiter et al., 2017; Wilke et al., 2017; Oña et al., 2019), I hypothesised that repertoire sizes will also be larger in contexts where the cost of ambiguity is particularly high; correspondingly, I predicted to find largest repertoire sizes in aggressive and playful contexts (e.g., threat of physical aggression). Building on this, based on the hypothesis that multimodal signals function to refine messages, I predicted that there will also be a high degree of context-specificity of different combination types, with the highest number of different combinations also occurring in contexts where message clarity is of increased importance. Examining the effect of interactional behaviour and context on multimodal combination frequency and diversity in this way can provide insight into the social pressures which may have encouraged the selection of increased multimodality in hominids ontogenetically as well as phylogenetically.

121

Methods

Study sites and subjects

The cross-sectional sample for this study is the same as that described in Chapter 2.

Data collection

Full details of the data collection procedure for this study are consistent with those provided in Chapter 2.

Behavioural coding

A total of 808 communicative interactions were observed across 394 video recordings. For each communicative interaction, all observed occurrences of facial expressions, vocalizations and gestures produced by the focal were coded using ELAN (version 6.0) open-source the video annotation software (https://archive.mpi.nl/tla/elan). Behavioural definitions of all signals coded within each communicative modality (facial expressions, vocalizations and gestures) are provided in Appendix A1.2.

As I am focused on the production of overlapping signal modalities, for every communicative interaction I counted the number of multimodal signal combinations produced (i.e., temporally overlapping distinct communicative acts). This number was '0' if all signals produced within an interaction (i.e., gestures, vocalizations or facial expressions) were produced singularly. For each communicative interaction, in addition to the communicative modalities observed I also coded: behavioural context i.e., feeding, play, aggression, grooming, travel, affiliation, access, rest (for definitions see Chapter 2); and (2) interaction partner ID of conspecific.

Inter-coder reliability

To assess the reliability of video coding, a second and third independent researcher also coded all signal events across 15% of the total number of video recordings (58 focal recordings, 27 individuals). Cohen's kappa (Cohen, 1960) was calculated for the reliable identification of each modality independently across coded signal events as well as if a signal event contained just one or multiple modalities simultaneously. The mean kappa value obtained for each modality indicated excellent levels of coder agreement (Fleiss, 1981) (facial expressions = 0.83; vocalizations=0.92, gestures=0.81). The level of agreement regarding the singular or simultaneous production of signal modalities as multimodal signal combinations was also excellent (0.81).

Statistical analyses

I investigated to what extent multimodal signalling development is influenced by mean communicative interaction rates and number of communication partners. Multimodal signalling development was measured via two matrices: (1) the frequency of multimodal combinations produced within communicative interactions and (2) the size of multimodal combination repertoires. For my predictor variables, mean interaction rate refers to the number of communicative interactions an individual was observed to participate in, divided by the total recorded-observation time for that individual; the number of communication partners refers to the number of different conspecifics an individual was observed to communicate with across the entire observation period. Across each predictor variable, I excluded interactions where the recipient was an individual's mother. This was for two reasons: first, because my primary interest here is in the role of the wider communicative environment in shaping multimodal communication but also, the interactive role of the mother would be inconsistent across my sample age range due to

varying degrees of spatial and behavioural independence throughout the immature period

(Goodall, 1986; Pusey, 1990). Additionally, to expand on the previous findings that

multimodal combinations were more frequently produced in the contexts of play and

aggression, I also examined the effect of interactional context on individual multimodal

repertoires.

Frequency of multimodal signals: I used a generalized linear mixed model to

examine the influence of the test predictors on multimodal signal frequency (GLMM:

Baayen, 2008). Multimodal signal frequency was based on the number of multimodal

combinations per communicative interaction. This included interactions where the

frequency of signal combinations 0, i.e., interactions wherein gestures, vocalizations

and facial expressions were produced singularly. As a result, the data contained an

excess of zeros due to a high number of communicative interactions where no

combinations were produced. To account for this, I constructed a zero-inflated

negative binomial model using the glmmTMB function of the R package 'glmmTMB'

(Brooks et al., 2017). I included mean interaction rate and total number of

communicative interaction partners as test predictors while controlling for age (in

years, range 1-11), sex (two levels: male, female), context (four levels: feeding, play,

aggression and 'other' see Chapter 2), group number (three levels: 1, 2, 4) and

observation year (two levels: 2017, 2021). As random effects (intercepts), I included

signaller and recipient identity including all relevant random slopes to keep type 1

error rates at the nominal level of 5 % (Schielzeth & Forstmeier, 2009, Barr et al.,

2013). To account for potential variation in the influence of communicative exposure

at different stages of ontogeny, I initially included interaction terms between age and:

interaction rate and total interaction partners in the full model. Information regarding

model checks and implementation procedures is consistent with those described in

Chapter 2. Model diagnostics are provided in Appendix 3.

Multimodal repertoire size: Utilizing a comprehensive repertoire of multimodal

signal combinations previously collected for this sample of chimpanzees (see Chapter 2), I

analysed whether multimodal repertoire sizes varied across ages, as well as whether they

were influenced by individual interaction rates, number of interaction partners and

interactive context. First, I calculated the repertoire size per individual (i.e., the number of

different multimodal signal combinations that the individual produced at least once).

Importantly, I make no claim that observed multimodal repertoires represent complete

repertoires across individuals. Rather, I used this variable as a proxy for the

communicative spectrum of individuals and without assuming that repertoire asymptotes

were reached. I used a generalized linear model (GLM; McCullagh & Nelder, 1989) with

a Poisson error structure and log-link function to analyse whether individual multimodal

repertoire size ($N = 28$) is associated with mean interaction rate and number of

communication partners. I again initially included interaction terms between age and

communicative interaction variables, which were removed if insignificant. All data was

processed in R v. 2.15 (The R Foundation for 461 Statistical Computing, Vienna, Austria,

http://www.r-project.org).

Lastly, I originally intended to analyse whether the proportions of multimodal

combination types used in one or several contexts differed statistically from a uniform

distribution using a Binomial test. The binomial test can determine whether a multimodal

combination type was only observed in one or more context as a result of chance. Chance

occurrence (typically represented by a proportion of 0.5) would mean that a single

multimodal combination type was just as likely to occur in one context as it is in >1,

statistically speaking. However, because of the large variability in multimodal

combination types and many of these only being observed once or twice in total (see

Chapter 2), I was unable to conduct the planned binomial test due to lack of statistical

power from cases with too few observations. Descriptive results are still presented.

Results

Multimodal combination frequency

Overall, I recorded a total of 332 multimodal combinations (individual mean ± SD: 12.52

± 9.05) from 808 communicative interactions across 28 immature chimpanzees (0-11

years old). Of the total number, 104 multimodal combination cases were produced by

infants (10.90 ± 9.15), 123 were produced by juveniles (18.42 ± 10.23) and 122 by early

adolescents (11.18 ± 7.13).

I tested the effect of mean interaction rate, number of interaction partners and

interactive context, on the production of multimodal signal combinations. The model

contained 769 data points, corresponding to the number of communicative interactions

for which the number of signals produced, and the identity of a recipient was clear.

Overall, the full model explained a significant amount of variation in multimodal signal

production compared to the null model (likelihood ratio tests comparing null and full

model (LRT): X^2 12.852, d.f. = 4, $P = 0.012$).

I did not find a significant interaction between age and mean interaction rate

(estimate ± se: -0.075 ± 0.093, $\chi^2 = 0.653$ d.f. = 1, $P = 0.419$) or age and total number of

communication partners (0.030 ± 0.095, $\chi^2 = 0.101$ d.f. = 1, $P = 0.751$) so both

interaction terms were removed from the model. The reduced model revealed that the

number of communication partners positively predicted multimodal combination

production (0.267 ± 0.125, $\chi^2 = 4.739$, d.f. = 1, $P = 0.030$) (Figure 4.1). That is,

individuals who interacted with a higher number of conspecifics produced multimodal

combinations more frequently. I did not find a significant effect mean interaction rate on

multimodal combination frequency (- 0.173 ± 0.141, χ^2 = 1.514, d.f. = 1, P = 0.219).

Lastly consistent with previous models (Chapter 2), multimodal signals were produced

more in play (0.833 ± 0.233, χ^2 = 14.867, d.f. = 1, P <0.001) and aggression (0.945 ±

0.259, χ^2 = 13.157, d.f. = 1, P <0.001) contexts than the 'other' reference category.

Effects of non-significant key predictors and other control variables are provided in

Appendix 3 Table A3.1.
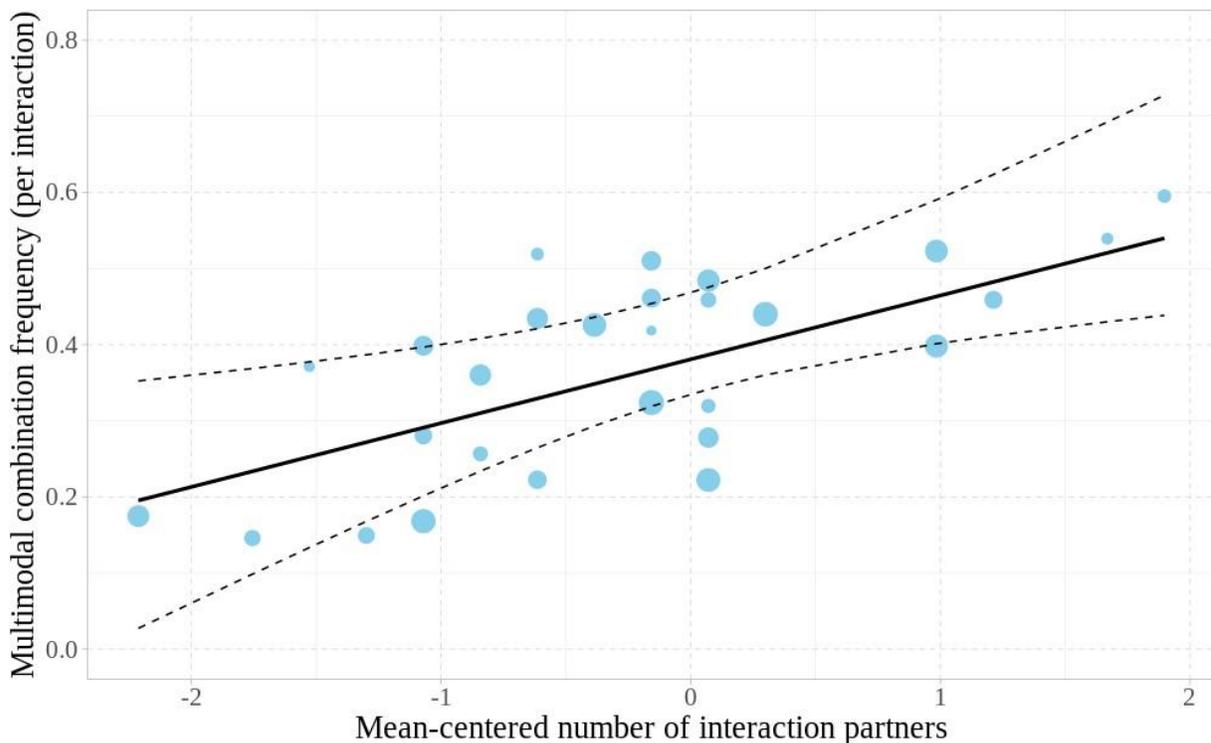
Figure 4.1. Effect of mean-centered number of interaction partners on the frequency of

multimodal signal production in immature chimpanzees. Depicted is the mean frequency of

multimodal combinations per communicative interaction across individuals (dots). The area of

the dots corresponds to the sample size per individual. The solid line and dashed lines

represent the fitted GLMM and 95% confidence intervals GLMM and 95% confidence

intervals, respectively.

Multimodal combination repertoire size

In total, I observed 110 different multimodal combination types of facial, gestural and vocal signals (Chapter 2). The largest multimodal repertoires were observed among juveniles (individual mean ± SD: 10.71 ± 5.44) followed by early adolescents (7.11 ± 5.15) and infants (5.70 ± 5.40). The combination 'open mouth + laugh' was observed most often, accounting for 21.48% of total multimodal signals recorded (N=78 occurrences).

Across all individuals, multimodal repertoire size was significantly associated with higher number of communication partners (estimate ± s.e. = 0.040 ± 0.026, t = 2.498, P = 0.021) (Figure 4.2) but not age (0.028 ± 0.036, t = 0.785, P = 0.442) nor mean interaction rate (0.040 ± 0.026, t = 1.563, P = 0.134). I did not find a significant interaction between age and either interaction variable. Effects of non-significant key predictors and other control variables are provided in Appendix 3 Table A3.2.
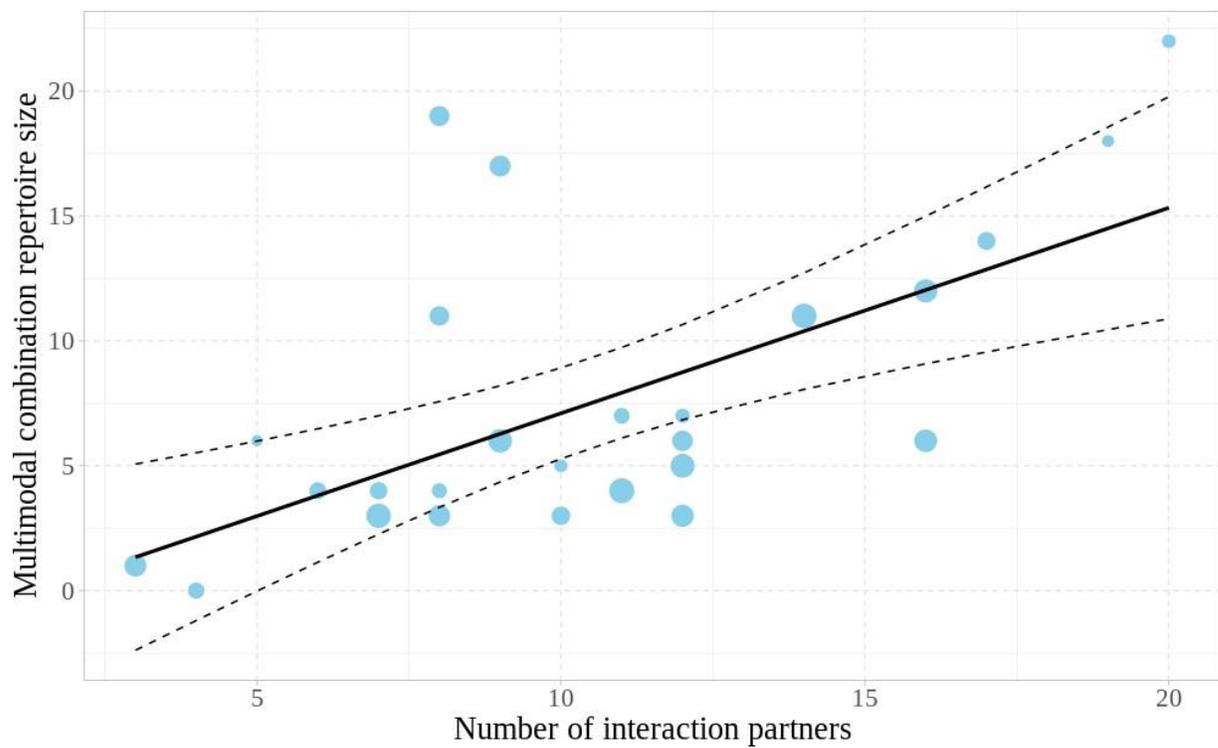
Figure 4.2. Effect of number of interaction partners on individual multimodal repertoire

size of immature chimpanzees. Depicted is the multimodal repertoire size per individual (dots). The area of the dots corresponds to the sample size per individual. The solid line and dashed lines represent the fitted GLM and 95% confidence intervals, respectively.

Context-specificity of multimodal combinations

Multimodal combinations were observed across all 9 interaction contexts. Most of the 110 combination types were used in the play (32.31%), aggressive (30%) and feeding interactions (16.15%) with the remaining number (37.69%) spread across the remaining 6 contexts (Figure 4.3a). Results indicated high context-specificity of multimodal signals. On average, each combination type was used in just one context (mean N of contexts per combination type $\pm$ SD = 1.17 $\pm$ 0.43). While 85.15% of combination types of the full observed repertoire were observed in a single context, only a small proportion were used in more than one (14.85%) (Figure 4.3b).
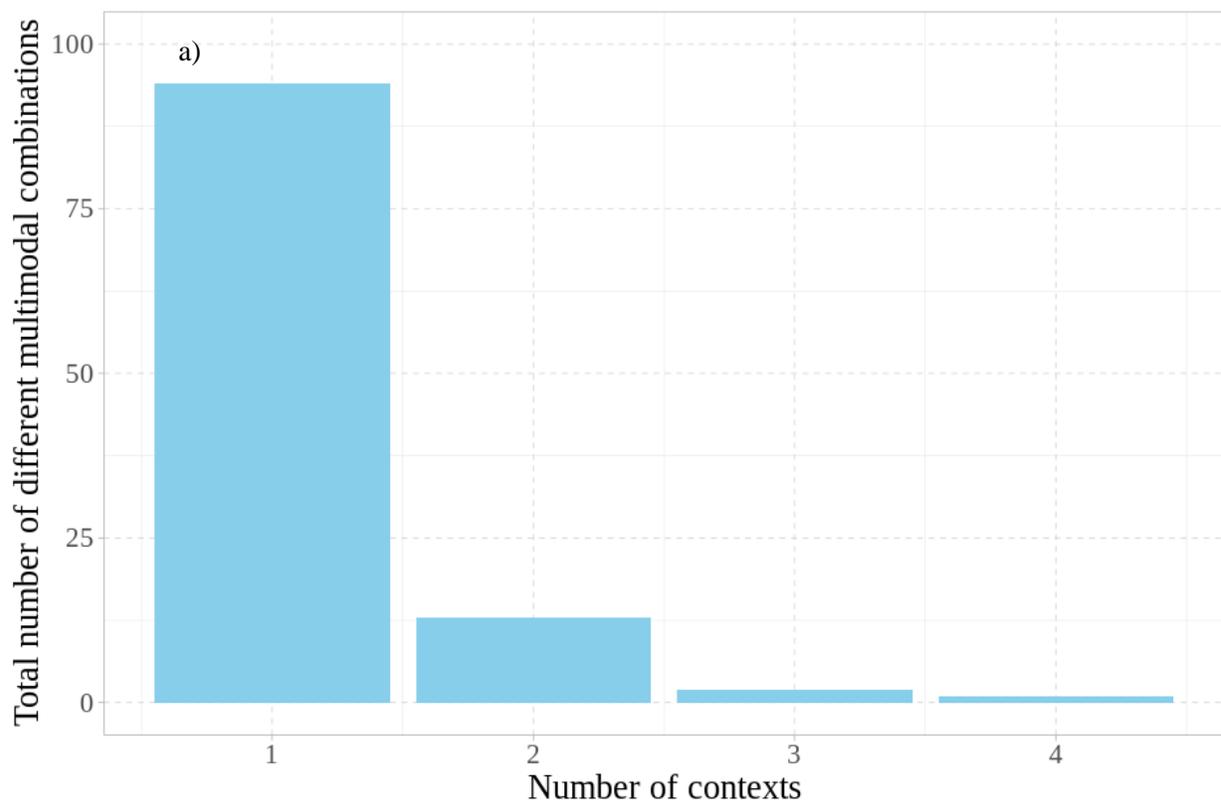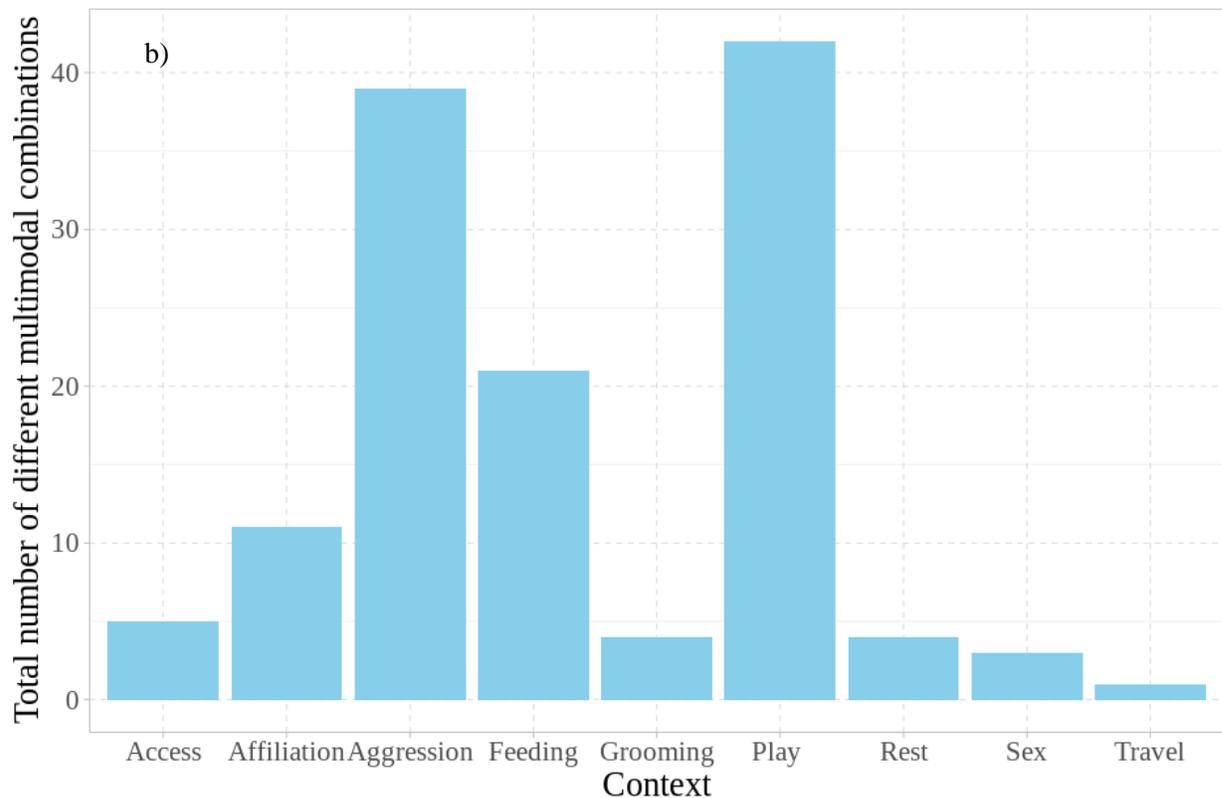
Figure 4.3. Contextual distribution of multimodal repertoire in immature chimpanzees (a) number of combination types as a function of the number of contexts in which they were recorded (b) number of combination types recorded in each context

Discussion

In this study, I examined whether factors related to observed (non-maternal) communicative interaction behaviour, namely interaction rate and number of communication partners, influenced multimodal combination production at different stages of chimpanzee development. To do this, I focused on two parameters of multimodal signalling: the frequency of multimodal combination production and multimodal repertoire size. Additionally, to expand on my previous findings that immature semi-wild chimpanzees produced multimodal combinations most frequently in playful and aggressive contexts (Chapter 2), I also investigated the role of interactional context on individual multimodal

combination repertoires and the degree of context-specificity in combination signals. I found that overall, a wider range of non-maternal communicative partners positively predicted both multimodal combination frequency and repertoire size regardless of age. However, in contrast to expectation, older individuals with higher mean interaction rates did not produce multimodal combinations more frequently; moreover, neither age nor interaction rate predicted multimodal repertoire size. Contextually, play and aggression stimulated the largest range of different multimodal combination types. Finally, I found a striking degree of context specificity of combinations in general. Overall, although some of my hypotheses were not met, the findings provide the first indication that social interactional environment influences multimodal signalling in immature chimpanzees.

While the first and primary interactional partner in a young chimpanzee life is their mother (Plooij, 1978;1979; (Chapter 3), developmental shifts from action-based to more actively communicative signalling occur at approximately 9-12 months (Plooij, 1987).

Coinciding increases in motor competence (Bründl et al., 2021) also mean that in comparison to young humans, young chimpanzees have more opportunity to engage in greater spatial independence and more frequent interactions with other social partners outside of the mother-infant dyad (Plooij, 1978; Fröhlich et al., 2016). These experiences within the wider social community are essential for the development of social competence and have long lasting effects on the social cognition skills involved in communication (Bard & Leavens, 2014). Here, my findings expand on research demonstrating the value of these interactive experiences on gestural (Fröhlich et al., 2017) and vocal (Taylor et al., 2022) development. Taking a multimodal view, they suggest that interactions with a wider number of communicative partners during the immature period also positively affects the production of

more complex communication. Although causal inferences cannot be made from a correlation, thus caution must be maintained, this observation appears to resonate with research in human children, which has consistently demonstrated that engaging with peers (Bates, 1975; Ervin-Tripp, 1991) and adults in child-care settings (McCartney, 1984; NICHD Early Child Care Network, 2000; Peisner-Feinberg et al., 2001) can have a positive influence on language acquisition. While I did not find a significant effect of age and thus no equivalent developmental pattern here, the findings overall could indicate that exposure to a wider communicative community influences communication skills in both young great-apes and humans. Given a wider range of communication partners, individuals may have the opportunity to produce and respond to a wider range of signal types, in contexts outside of those experienced in interactions with the mother alone e.g., peer social-play or competition-based aggression. This exposure could in turn encourage both the learning of and need to use clarifying multimodal combinations (Pusey, 1990) which would explain the positive effect I observed on the frequency and diversity of multimodal combinations used across communicative interactions.

In contrast to my initial prediction, I found no effect of mean interaction rate on multimodal combination frequency or repertoire size, nor any interaction between age and interactive variables on either metric. While I am not aware of studies tracking the effect of interaction rates on the production of vocal or gestural signals, this finding contrasts with Fröhlich et al.'s (2017) study of gestural development wherein with increasing age, higher mean interaction rates positively affected both gesture frequency and production in sequences, but not repertoire size. Considering the finding that more communication partners did have a positive effect on multimodal combination use, it could be that it is more important to interact with a wider range of conspecifics rather than interacting more in general.

Another possibility is that this lack of finding could reflect more of a context-dependent effect on these signals, wherein multimodal combination production is more related to the type of interaction in which they are produced, for example where their disambiguating function is most advantageous e.g., play and aggression contexts. This is consistent with the strong effect of context on combination repertoires found here. However, a third, more parsimonious explanation could be that my cross-sectional sample restricted my ability to capture patterns of interactive behaviour occurring across time representing a limitation of this study. To test these hypotheses, longitudinal data of individual interactional experience is required. This would allow us to evaluate potential learning mechanisms of multimodal combinations more fully discriminate from, or relate these, to a more context-induced role of production. One particularly fruitful avenue will be to explore the effectiveness of multimodal combinations across time to understand the feedback young chimpanzees receive within interactions which may contribute to their production in future interactions.

A notable finding of this study relates to the degree of different combination variations contained within contexts. As predicted, I found that the largest amount of variation within individual repertoires was observed within playful and aggressive interactions. So, not only are multimodal combinations used more frequently within these contexts, they also are much more diverse in communicative structure. Given the highly interactive nature of play, such a finding is perhaps not surprising but nevertheless highlights the importance of play for socio-communicative development (see for review: Palagi, 2018). In my analysis, the result was largely explained by the combination of the 'open mouth' play expression with different gestures and/ or laughter. This expression is an important an important social regulator in play

133

interactions (Waller & Dunbar, 2005; Davila-Ross & Dezecache, 2021) with well-known

multimodal links to 'laughter' vocalizations (Davila-Ross et al., 2015) and also appears

to function to disambiguate the intention of rougher gestures like hitting or grabbing

clear as playful rather than aggressive (Davila-Ross & Palagi, 2022). Similarly, the also

large variation observed across aggressive interactions is consistent with a 'clarifying'

function of multimodal combinations (Pollick & de Waal, 2007; Genty et al., 2014;

2015b, Wilke et al., 2017, Genty, 2019; Oña et al., 2019), in this case where the costs of

miscommunication are particularly high e.g., physical aggression. The high degree of

context-specificity of multimodal combinations found here highlights the importance of

multimodal perspectives to the investigation of comparative properties of human and

non-human communication systems. Studies of chimpanzee gestures have cited their

flexible production across contexts where meaning is extracted from the context in which

they are produced (Hobaiter & Byrne, 2011a, 2014) while facial and vocal signals have

more traditionally been viewed as bound to specific functions (see for review: Liebal &

Oña, 2018). However, while the production of single communicative components may

appear flexible, the use in combination with additional signals in fact appears to be much

more contextually bound. Without considering additional signals that accompany the

flexible, unimodal components, important patterns may be missed which may be leading

to a misunderstanding of the mechanisms underlying communication systems and their

acquisition.

Moving forward, further investigations of functional specificity in multimodal

combinations between and within social interactions will be extremely valuable. Here,

although I found a large number of multimodal combinations, many of these different

combinations were observed in just one context. This could be indicative of a high degree of

combinatorial flexibility across different signal modalities, however due to many combination types only being observed once or twice in my dataset, I was unable to test this statistically due to a limited predictive value of my findings. While many combination types are intrinsically linked to context given the type of signals contained within the combination (e.g., open mouth (play) faces in the play context, or screams in the context of aggression) more data is required to assess the true degree of context specificity of multimodal combinations. In addition, an ongoing challenge surrounds the adequacy of 'context' coding in primate behavioural research which may lead to important interactional information being missed by imposing human judgements about distinction and relevance of different behavioural contexts (see General Discussion). Like our own communication, the exact context of an interaction is likely much more nuanced, which incorporates aspects including the signallers affective state and behaviour and immediate social environment, as well as factors related to a particular recipient such as their age, familiarity and/ or relative rank (Graham et al., 2020). I recognize that in the absence of more fine-grained analysis of context, I may not be reporting the true degree of 'context' specify of multimodal combinations. While this remains a complex issue to resolve, another fascinating line of investigation would be to consider components of multimodal combinations may bound to a particular function (e.g., crouch posture is an indication of submission), and then examine how different, additional components modify this function across contexts/ recipients. Here, it would be valuable to incorporate a greater emphasis on social network matrices, such as estimates of network size, diversity and egalitarian structure. This would provide more subtle insight into the interactions between the complex social environment of chimpanzees and its relationship with forms of complex communication such as multimodal signalling which would have important implications

for our understanding of the social-selective drivers of multimodal language.

Another potential limitation of this study is the lack of control for observation time when including the number of interaction partners and multimodal repertoire sizes of each individual in the analysis. I acknowledge that time observed could have influenced these matrices across individuals and therefore be affecting my findings. I would encourage such consideration in future analyses by including observation time in the construction, and interpretation of models.

Lastly, another important consideration in the interpretation of my current findings is the research setting from which data was collected, which may have particularly affected the contexts in which I observed communication most. Although all chimpanzees in my study lived in large, forested enclosures, my observations largely took place during periods surrounding twice-daily artificial food provisioning when focal subjects were visible near the enclosure fence lines. The build up to, and occurrence of food-provisioning is likely to enhance social tensions within the group and by extension, social behaviours associated aggression and tension mitigation e.g., play (Palagi et al., 2004). While it is still informative that multimodal communication seems to be particularly associated with these 'higher stakes' contexts, I acknowledge that my observations may not necessarily reflect patterns of behaviour in wild populations or across other research settings. The continued investigation of multimodal signal combinations across different ape populations and contexts at an interaction level is vital to elucidate communicative patterns and function of multimodality as well as its evolution.

Conclusions

The present study examined the influence of patterns of communicative interactive behaviour on the development of multimodal combination signalling in immature chimpanzees. Overall, I demonstrated that communicative interactions, specifically the number of individuals a young chimpanzee interacts with, has a positive effect on the frequency and repertoire size of multimodal signal combinations. Interacting with a wider range of conspecifics could aid in the learning of new signals outside of the mother-infant dyad and drive appropriate use in social contexts. Moreover, I have provided novel insight into the contextual-specificity of multimodal combinations which highlights the importance of multimodal perspectives when considering the flexibility of great ape signals. Contrary to my prediction, I did not find age-effects of interaction behaviour on combination production, which may represent a limitation of my cross-sectional sample. While this study provides some evidence that interactional environment positively effects multimodal communication, future studies should focus on longitudinal investigations of chimpanzees living in their natural environment and incorporate more fine-grained analyses of social interactions. In turn, this will shed light on the influence of complex social pressures on the evolution of multimodal language.

Chapter 5 - The developing role of multimodality in the communicative interactions of immature chimpanzees

Abstract

Research has revealed that like humans, other great apes also flexibly combine facial, gestural, and vocal signals in their daily communication which aids in the disambiguation of messages. However, studying the use of multimodal signal combinations (or multimodal combinations) across ape ontogeny is needed to shed light on its cognitive precursors, and the selective pressures which encouraged multimodality in hominins. Here, I studied the communicative interactions of a cross-section of N=28 sanctuary-living chimpanzees aged 1-11 years to test predictions based on the hypothesis that multimodal combinations have 'disambiguation' function in ape communication. Specifically, I examined age-related variation in the use of, and responses to, multimodal combinations which could indicate their ability to enhance effectiveness and achieve communicative goals. In support of the hypothesis, our key findings were that older individuals produced more multimodal combinations when a recipient was visually oriented towards them, indicating an increasing degree of understanding of recipient's attentional state, suggestive of possible intentionality. Additionally, after controlling for context and interaction partner, signal events containing multimodal combinations (regardless of age) were more likely to promote responses than those which contained unimodal signals only, and these responses likely to satisfy the apparent communicative goal of the signaller. These results indicate that similar to humans, even young chimpanzees may have the capacity to use multimodal combinations to enhance effectiveness in pursuit of a communicative goal.

Introduction

Human language is an inherently multimodal communication system, wherein speech or sign is routinely integrated with visual information extracted from eye gaze, facial expressions, gestures and bodily postures, acting to complement and refine the signaller's intended message (e.g., Goldin-Meadow, 1999; McNeill, 2000; Kendon, 2004; Holle & Gunter, 2007). The coordination of distinct communicative modalities is present even in very early infancy (Yale et al., 1993; 2003) and as development proceeds, this ability continues to become strengthened and refined as children become capable of more intentionally communicating meaning via multimodal expressions (Bates et al., 1975, 1979; Wetherby & Prizant, 1989; Liszkowski et al., 2008; Igualada et al., 2015). Furthermore, the production of effective, multimodal combinations, such as gesture-word combinations, has been shown to promote later language abilities in humans (Murillo & Belinchón, 2012; Murillo et al., 2018) and as such, appears to be a developmental precursor to language.

Investigating whether these capacities also underlie the production of multimodal combinations in our closest living relatives can therefore provide insight into the evolutionary origins of language.

Concerning the simultaneous coordination of distinct communicative acts (i.e., multimodal signal combinations (Fröhlich & Hobaiter, 2018; Fröhlich et al. 2019; Chapter 2), it has been established that our closest great ape relatives also use a substantial repertoire of facial, gestural and vocal combinations (e.g., Wilke et al., 2017; Chapter 2). At the proximate level, new evidence from adult chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*) has indicated like humans, (non-human) ape multimodal signal

combinations (henceforth also referred to as multimodal combinations) may function to aid comprehension by disambiguating a core message. For example, recent work has shown that semi-captive bonobos use different gestures to clarify playful or aggressive intent of the 'contest-hoot' vocalization (Genty et al., 2014). In wild chimpanzees, Wilke et al. (2017) reported that bimodal signals (gesture + 'grunt' vocalization) were more likely to elicit behavioural responses than when the vocalization was used alone, but not when the gesture was used alone. Furthermore, while inclusions of facial expressions are rare, Oña et al., (2019) also presented evidence that when combined with a particular gesture, facial expression had an augmentative effect on recipient response in semi-wild chimpanzees, which were different than when the same gestures were produced alone.

Given its apparent role in great ape communication, is likely that combinatorial signalling as a form of message-clarification was already present in the communication systems of our most recent *Pan-Homo* ancestor. If so, this would suggest that multimodality could have been important in paving the way for language to subsequently evolve. However, multimodal research is still in its infancy with respect to great ape communication, and important gaps remain in our understanding of the role of ape signal combinations and associated cognitive requirements. Particularly, it is critical that ontogenetic perspectives are applied to this topic. The process of development is sensitive to modification from external pressures in a manner analogous to the evolution of behavioural endpoints (Brigandt & Love, 2010). Expanding on existing studies of multimodal communication in great apes by considering the effect of ontogeny on multimodal signal production within communicative interactions is needed to provide comparable insight into the underlying cognitive mechanisms and role of the socio-ecological environment in shaping multimodal communicative systems. While great ape

multimodal combinations appear to share a similar, intentionally clarifying function to human multimodal messages, we do not yet know when this skill emerges in ontogeny nor the factors which contribute towards its acquisition. Studying the circumstances of production of and responses to multimodal signal combinations at different stages of ape ontogeny is critical to understand the proximate and ultimate mechanisms driving their manifestation and functional role in adult repertoires.

The main aim of the present study was thus to investigate the clarifying role of multimodal combinations within the communicative interactions of immature chimpanzees at various stages of ontogeny. To do so, I observed the communicative behaviour of 28 sanctuary-living individuals ranging from early infancy to adolescence. I examined factors influencing both the use of multimodal signal combinations by young chimpanzees in my sample, and how these combinations effected recipient response behaviour and interactive outcomes. Specifically, I focused my attention on three research questions: First, regarding multimodal combination production: Does the identity and visual attention of the recipient influence the likelihood of multimodal signal production? To address this question, I examined whether recipient was visually oriented towards the signaller or not at time of signal production; and distinguished recipient relationship between mother, maternal kin and non-kin (i.e., neither mother nor maternal kin). In humans, experimental studies of joint attention initiation in infants have shown that infants more often use vocal-gesture combinations when an adult is visually oriented towards them than when they are not (Gros- Louis & Wu 2012; Wu & Gros-Louis, 2014; Liszkowski et al., 2008; Igualada et al., 2015). This likely reflects an ability to intentionally use simultaneous signal combinations to reinforce information related to their communicative goal when an adult does not share attention to a referent but is

crucially available to communicate. Based on the hypothesis that multimodal combinations have disambiguation in ape communication, I further hypothesized that immature chimpanzees will use multimodal combinations most often when there is a higher chance that they can be understood. I therefore predicted that they will be produced most often when the signaller is visually orientated towards the signaller so all information contained within a signal can be relayed. Furthermore, a clarifying function would mean that signal combinations may be particularly relevant when social outcomes are less predictable due to lower degrees of familiarity and social tolerance. Recent work on Orangutans (Fröhlich et al., 2021) has supported this supposition, showing that signal combinations were much less likely between mother-infant dyads (high social tolerance) than other interactive dyads. Thus, I also predict that signal events will contain multimodal combinations most often when the recipient is less familiar to the signaller e.g., non-kin (i.e., lower outcome predictability).

Second, do multimodal signals affect the probability, latency and type of recipient responses? In line with findings from previous great ape research where facial, gestural and vocal signals have all been considered (Fröhlich et al., 2021), I predict that signal events containing multimodal combinations will be more effective in soliciting recipient responses than those which do not. Again, based on the hypothesis that these signals function to add clarity to ape communicative messages, not only would I predict responses to be more likely, but multimodal combinations will also promote faster, and more positive responses.

Third, are multimodal signals more successful in achieving a com municative goal? To answer this question, I used the method of goal-outcome matching (Cartmill &

Byrne, 2010), whereby I assigned an apparent aim of the signaller to each signal, based on the individuals involved and the immediate social context in which the interaction occurs. Communicative success was then attributed to signals where the goal of the signaller seems satisfied by the interaction outcome e.g., food begging results in food sharing. Based on the hypothesis that chimpanzees produce multimodal combinations to disambiguate communicative messages, I predict that that combinations will promote successful interaction outcomes, as indicated by the achievement of apparent communicative goals (Cartmill & Byrne, 2010). Lastly, in addition to all predictions outlined here, I also predict age effects related to each research question. Specifically, according to the hypothesis that chimpanzees learn to communicate more effectively over time, I predict that each of the predicted effects will be stronger in older individuals due to development of socio-cognitive skills.

Methods

Study sites and subjects

The cross-sectional sample for this study is the same as that described in Chapter 2.

Data collection

Full details of the data collection procedure for this study are consistent with those provided in Chapter 2.

Behavioural coding

To assess the factors contributing to the production of, and responses to multimodal signal combinations, I focused on signal events wherein a focal individual initiates or reinitiates a communicative interaction with a clear recipient, as indicated by directed eye gaze or physical touch. I define a signal event as the production of a single

communicative act, but

also a string of communicative acts if they are produced <1sec apart. A communicative

act is defined as an individual facial expression, gesture, vocalization or a temporally

overlapping combination of these (multimodal signal combination). Communicative acts

produced >1s apart were treated as separate signals but included within a single signal

event. This resulted in a total of 598 signal events being recorded across 476

communicative interactions.

All signal events were coded using ELAN (version 6.0), an open-source video

annotation software (https://archive.mpi.nl/tla/elan). Behavioural definitions of distinct

communicative acts contained within signal events were based on established ethograms

for chimpanzees (facial expressions: Parr et al., 2005, 2007; Bard et al., 2011;

vocalizations: Plooij, 1984; Goodall, 1986; Kojima, 2008; Slocombe & Zuberbühler,

2010; gestures: Nishida et al., 1999; Byrne et al., 2017). Full descriptions of all coded

behaviours are given in Chapter 2 and Appendix 1.3 Tables A1.4, A1.5 and A1.6.

In addition to communicative act type, for each signal event I specified: whether

it contained a multimodal signal combination (yes, no), the identity of the recipient, kin

relationship to signaller (mother, maternal kin, non-kin), whether or not the recipient was

visually oriented towards the signaller at the *beginning* of signal production (visually

oriented: focal is in the recipients full to mid-peripheral view i.e., within an arc of 60°

either side of direction recipient is facing; visually unavailable: focal is outside of the

recipient's mid-peripheral view). If a recipient turned to look at the signaller during a

signalling event, this was still recorded as 'visually unavailable' given that this was this

case when the focal commenced signalling. I coded signals within nine behavioural

contexts (see Chapter 2 for context descriptions) but as the vast majority focal-initiated signals occurred in play and feeding context, I only included these two contexts in our models. The remaining contexts were grouped as 'other' context category and used as a reference for feeding and play.

Regarding recipient response behaviour, for consistency with recent research I followed Wilke and colleagues (2017) by coding recipient responses as from the beginning of a focal-produced signal until 20s after it ended. If the focal produced another signal (i.e.,
>1s after end of previous signal) within this time, any following response was coded for this signal only. I also coded the latency (in milliseconds) of responses from the beginning of the focal signal until this 20s cut off. Recipient responses were split into three groups: signal, positive behavioural and negative behavioural (Wilke et al., 2017) (see Appendix 4 Table A4.1). Signal responses corresponded to communicative behavioural definitions applied to focal subjects (e.g., facial expression, gestures, vocalizations, multimodal combinations) but were only counted if they could be reliably determined to be directed towards the focal individual. Cases where this could not be established were coded as 'unknown' and excluded from analysis.

Lastly, to assess whether signal events containing multimodal combinations were more successful in achieving apparent communicative goals, I adapted the method of 'goal- outcome matching' previously used to determine the meaning of specific signals in gestural research (Cartmill & Byrne, 2010). I considered the behaviour of both the focal and recipient from the beginning of a focal signal event to the outcome and considered the outcome 'successful' if it matched the apparent goal of the signaller (e.g., joint travel,

play with me, move away). For example, if a focal subject initiated food begging from a feeding recipient and the recipient subsequently shared a food item, this was coded as successful. Full descriptions of coded communicative goals are provided in Appendix 4 Table A4.2.

Intercoder reliability

This data used in this analysis is part of a larger dataset of communicative signals produced by the study sample (Chapter 2 and 3). Previously, a second and third independent researcher coded the occurrence of signals produced across 15% of the total number of video recordings in the original dataset (58 focal recordings, 27 individuals). Cohen's kappa (Cohen, 1960) was calculated for the reliable identification of each signal modality (facial expressions, gestures, vocalizations) independently across coded signal events as well as whether these signals were produced in isolation (unimodal) or in combination (multimodal signal combination). The mean kappa value (Fleiss, 1981) was 0.84 indicating excellent levels of coder agreement.

For the current analysis, a second independent researcher also coded 10% of total number of video recordings contained within the current dataset (29 focal recordings, 21 individuals). As the reliability of signal modalities contained within signal events was already accounted for, in this instance Cohen's kappa was calculated for the reliable identification of recipient's visual orientation at the time of signal production (yes/no, κ=0.96), recipient response (yes/no, κ=0.83) and communicative success (i.e., achievement of a communicative goal) (yes/no, κ=0.78). These values represent an excellent, and very good level of agreement, respectively.

Statistical analysis

To address the three research questions and investigate sources of recipient-related variation in multimodal signalling events (model 1); the effect of multimodal combinations on response probability, latency and type (model 2-4); and whether multimodal combinations are more successful in achieving apparent communicative goals (model 5) I used GLMM's (Baayen, 2008) with a Binomial (model 1,2,5) or Poisson (model 3,4) error structure and a logit (model 1,2,5) or log (model 3,4) link function. To allow full advantage to be taken of coded information, I used a modified dataset for each model corresponding to the response variable after rows with non-available cases for that response variable were omitted, in addition to non-available cases for each predictor.

In model (1), I included focal age (range: 1-11 years), whether a signal event contained a multimodal combination (yes, no), whether the recipient was visually oriented towards the signaller (yes, no) and kin relationship (mother, maternal in, non-kin) as key test predictors. In models (2-5), our key focus was the effect of multimodal combinations on recipient responses and interaction outcomes so here our key test predictor was only whether a signal event contained a multimodal combination (yes, no). In these models, visual orientation of recipient and kin relationship to signaller were still included, but as control predictors as they were not directly related to the questions being asked. In all models I also included focal sex (male, female), context (feeding, play, other), group (1, 2, 4) and observation year (2017, 2021) as control predictors. To account for the influence of developmental stage of the focal on the effect of key test predictors, in all models I also initially included interaction terms between focal age and

147

variables related to the recipient. All interaction terms were initially included in each full model, and significance was assessed. If insignificant, interaction terms were excluded from further analysis, and I ran a reduced model containing only the respective variables to allow for more accurate interpretation of fixed effects.

To control for repeated measurements, I included focal, recipient and dyad ID as random effects. Since communicative interactions could include signal events containing multimodal combinations and signal events which did not, I further included interaction ID as a random effect, to control for the fact that signal events within the same interaction are not independent. To keep type 1 error rates at the nominal level of 5%, I also included relevant random slope components within-subject and recipient ID (Schielzeth & Forstmeier, 2009). I did not include correlations between random slopes and random intercepts to keep model complexity at an acceptable level and because neglected random slopes do not compromise type 1 error rates (Barr et al., 2013). Information regarding model checks and implementation procedures is consistent with those described in Chapter 2. Diagnostics for all models are provided in Appendix 4.

Results

A total of 598 initiative signal events were coded during 476 communicative interactions across 28 focal individuals (individual mean ± SD: 21.25 ± 16.19). Altogether, 162 signal events involving 24 focal individuals across all age classes included multimodal combinations (5.46 ± 4.90). Visible recipient responses were recorded in response to 292 (48%) signal events (10.43 ± 9.83) of which 193 appeared to match the focal-signallers apparent communicative goal (6.89 ± 6.84). Importantly, these estimates do not include cases when a recipient's response behaviour was not visible, therefore the relative

proportion of responses which appeared to match a communicative goal may be either conservative or an overrepresentation of communicative success.

Does recipient visual orientation and identity influence the probability of multimodal combination production?

In my first model, I tested the effect of recipient visual orientation (at the time of focal-signal production) and kin relationship to the signaller and on the probability of multimodal combination production in immature chimpanzees. The model contained 564 data points, corresponding to the number of focal-initiated signals for which the ID and visual status of the recipient was available. Overall, the full model explained a significant amount of variation in the probability of multimodal signal production compared to the null model (likelihood ratio tests comparing null and full model (LRT): χ^2 =12.140, d.f. = 3, $P = 0.007$).

 After removing non-significant interactions, the reduced model revealed a significant effect of both focal age (estimate $\pm$ s.e = 0.571 $\pm$ 0.225, $\chi^2 = 6.651$, d.f. = 1, $P = 0.010$) and recipient visual orientation at the time of signal production (0.781 $\pm$ 0.375, $\chi^2 = 4.914$, d.f. = 1, $P = 0.027$). That is, older individuals were more likely to produce multimodal signal events than younger individuals and, regardless of age, multimodal combinations were more likely to be included when the recipient was visually oriented towards the signaller (Figure 5.1). I did not find effects of kinship (mother: -0.109 $\pm$ 0.438, $\chi^2 = 0.062$, $P = 0.802$; maternal kin: 0.121 $\pm$ 0.405, $\chi^2 = 0.072$, $P = 0.789$). Lastly, while not the focus of investigation signal events containing multimodal combinations were more frequent in the play context (1.186 $\pm$ 0.390, $\chi^2 = 8.316$, d.f. = 1, $P = 0.004$). Effects of non-significant test predictors and control variables are provided in Appendix 4 Table A4.3.
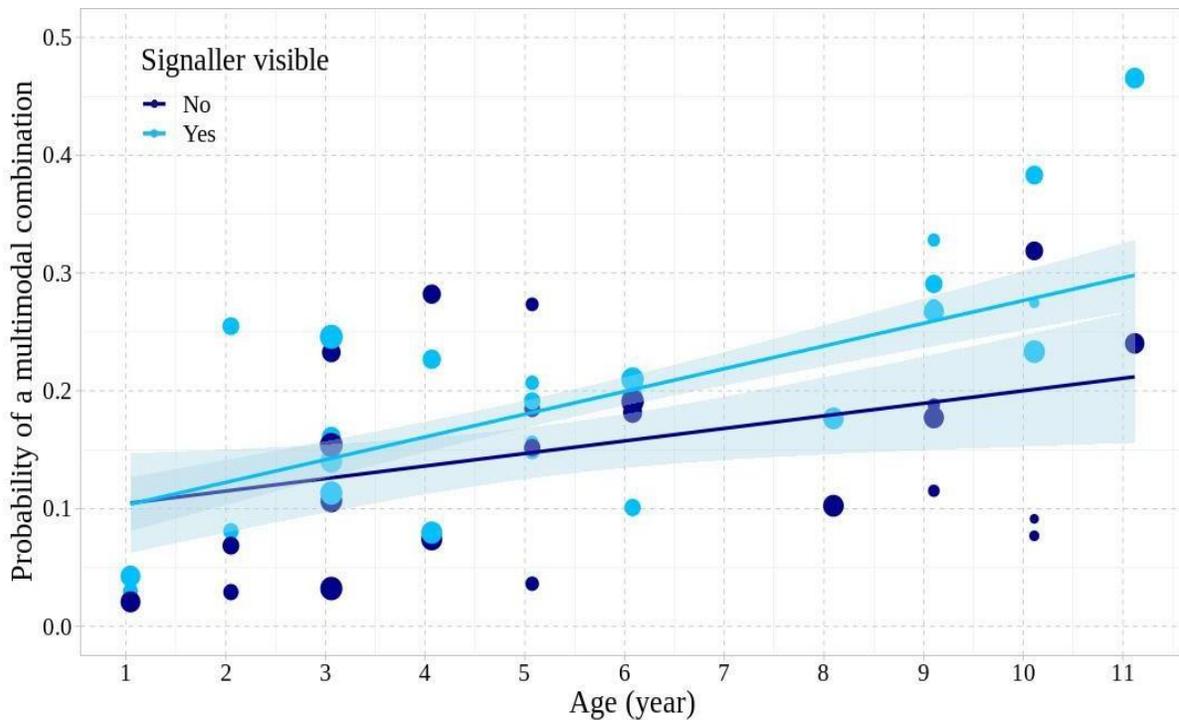
Figure 5.1. Effect of signaller visibility and signaller age on probability of a signal event containing a multimodal signal. Depicted are the mean number of multimodal events produced for each year of age. The size of the dots corresponds to the sample size per individual. The solid line and shaded area represent the fitted GLMM and 95% confidence intervals, respectively.

Do multimodal signals influence the probability, latencyand type of

recipient responses?

I used three models to test whether the effect of multimodal signals affected the probability, latency and type of recipient responses in immature chimpanzees. Each model contained 542, 237 and 277 data points, respectively, corresponding to the number of focal signals for which relevant response information was available. Overall, the test predictors had a clear impact on the probability and latency of responses (LRT for response probability: $\chi^2 = 6.019$, d.f. $= 2$, $P = 0.049$; response latency: $\chi^2 = 427.96$, d.f. $=$

150

3, *P*<0.001) but not response type (LRT for response type: $\chi^2 = 2.953$, d.f. = 2, *P* = 0.863).

Regarding response probability, once insignificant interactions were removed, the reduced model revealed that signal events containing multimodal combinations had a significant effect on the probability of receiving a response (0.568 ± 0.260, $\chi^2 = 4.893$, d.f. = 1, *P* = 0.027). That is, signal events containing multimodal combinations were significantly more likely to elicit a response, regardless of age (Figure 5.2). In addition, response probability was higher when the signaller was in view of the recipient (0.826 ± 0.280, $\chi^2 = 9.224$, d.f. = 1, *P* = 0.002) and when the recipient was the mother (1.282 ± 0.397, $\chi^2 = 11.725$, d.f. = 1, *P*<0.001). Lastly, although not the focus of this investigation, there was also a significantly higher probability of response in the context of play (1.437 ± 0.302, $\chi^2 = 19.671$, d.f. = 1, *P* <0.001), while the opposite was true in feeding-related interactions (-0.615 ± 0.426, $\chi^2 = 4.412$, d.f. = 1, *P* = 0.036). Effects of non-significant key predictors and other control variables are provided in Appendix 4 Table A4.4.
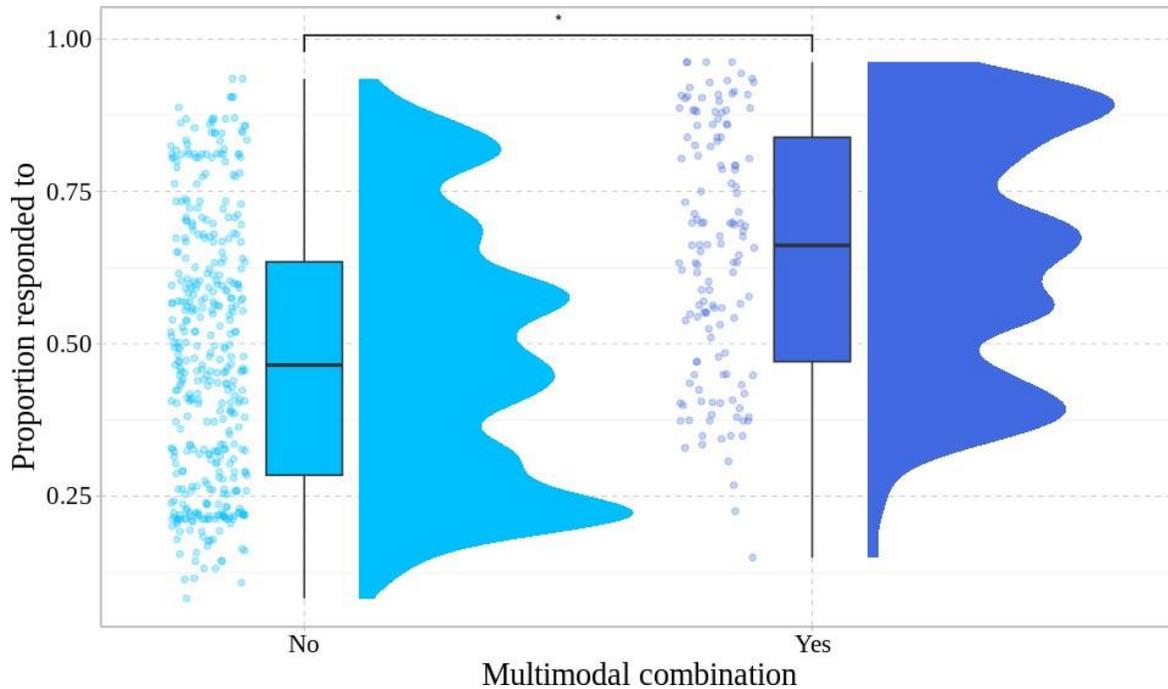
Figure 5.2. Effect of multimodal combinations on the probability of eliciting a recipient response. Indicated are communicative interactions (dots) the mean number per condition (white circle), median (horizontal lines), quartiles (boxes), percentiles (2.5 and 97.5%, vertical lines). Asterisks represent significant values (*=<0.05; **=<0.01; ***=<0.001).

For response latency, I found a significant, negative interaction between age and signal type on response times (-0.200 ± 0.028, $\chi^2 = 55.019$, d.f. $= 1$, $P < 0.001$). In older individuals, signal events containing multimodal signals were responded to significantly faster than those which did not (Figure 5.3). Effects of non-significant key predictors and control variables are provided in Appendix 4 Table A4.5.
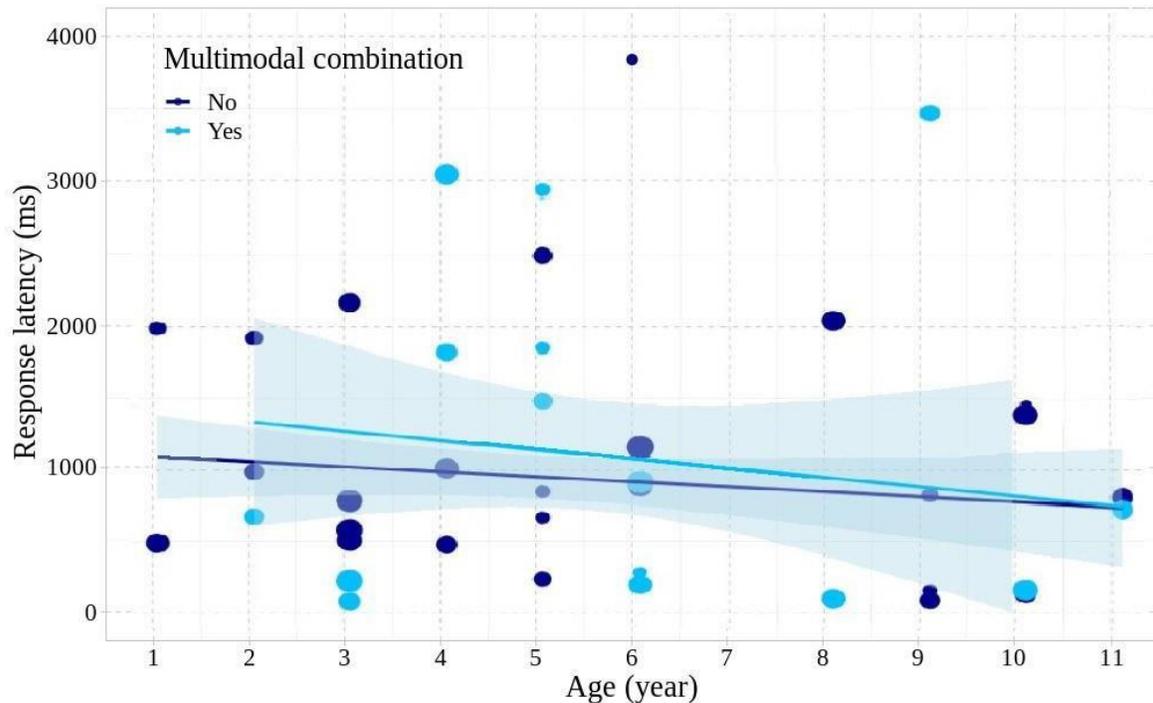
Figure 5.3. The effect of multimodal combinations on response latency [milliseconds] as a function of age [year]. Depicted are the mean response times for each year of age when a multimodal signal is absent or present. The size of the dots corresponds to the sample size per individual. The solid line and shaded area represent the fitted GLMM and 95% confidence intervals, respectively.

Do multimodal signals promote the reaching of an apparent communicative goal?

Finally, I tested whether the use of multimodal combinations predicted the probability of the signaller reaching an apparent communicative goal. The model contained 353 data points corresponding to the number of focal-initiated signal events here this could be determined. Overall, the full model fitted the data better than the null model (LRT: $\chi^2 = 7.539$, d.f. $= 2$, *P* $=0.023$).

After insignificant interactions were removed, results from the reduced model showed that the use of multimodal combinations in signal events produced by immature chimpanzees

had a significant positive effect on the probability of achieving an apparent communicative goal (0.868 ± 0.360, χ^2 = 6.215, d.f. = 1, P = 0.013). That is, multimodal signal events were significantly more likely to achieve an apparent goal across all ages (Figure 5.4). I also found that communicative goals were more likely to be achieved regardless of modality when the recipient was: visually oriented towards signaller (0.922 ± 0.3401 χ^2 = 5.809, d.f. = 1, P =0.016) and a signaller's mother (1.422 ± 0.498, χ^2 = 8.020, d.f. = 1, P = 0.005); but less likely in the feeding context (-2.268 ± 0.542, χ^2 = 23.002 , d.f. = 1, P <0.001). Effects of non-significant control variables are provided in Appendix 4 Table A4.6.
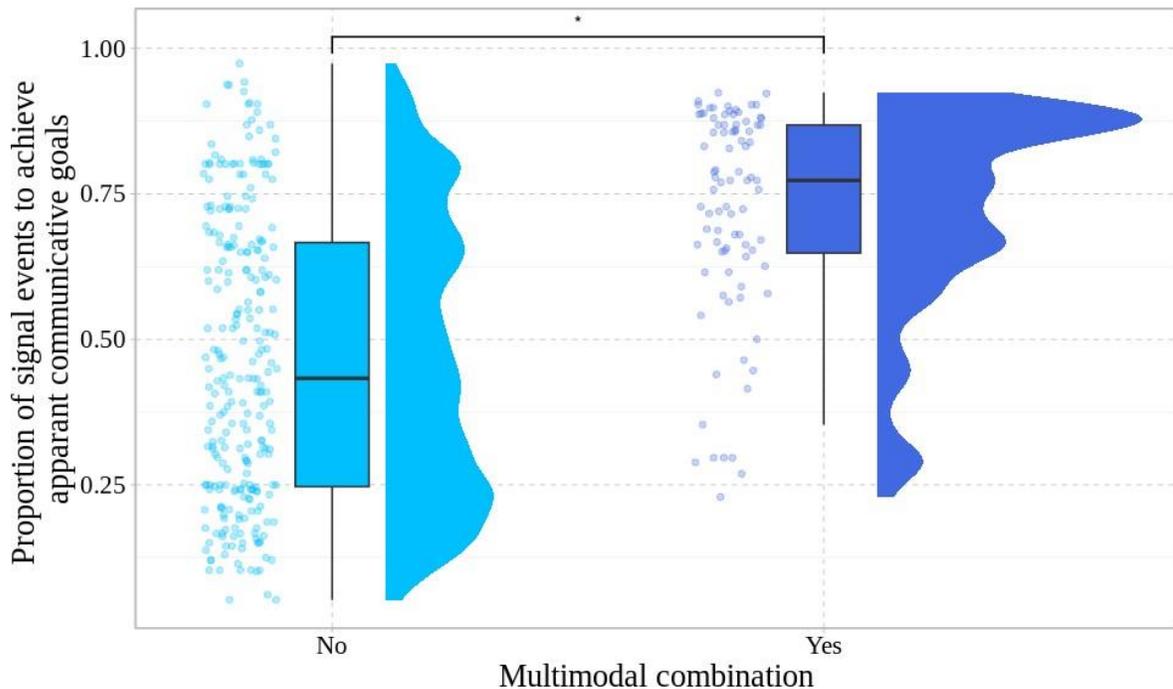
Figure 5.4. Effect of multimodal combinations on the probability of achieving an apparent communicative goal. Indicated are signal events (dots), median (horizontal lines), quartiles (boxes), percentiles (2.5 and 97.5%, vertical lines) and distribution of data (shaded areas). (*=<0.05; **=<0.01; ***=<0.001).

Discussion

The current study aimed to expand on the hypothesis proposed in the literature of a clarifying function of multimodal signal combinations in adult ape repertoires (e.g., Pollick & de Waal, 2007; Genty et al., 2014; 2015b, Wilke et al., 2017; Oña et al., 2019) by investigating the proximate factors contributing to their production at different stages of ontogeny. To do so, I observed the communicative behaviour of 28 sanctuary-living chimpanzees ranging from early infancy to adolescence. I examined factors influencing both the use of multimodal combinations by young chimpanzees in my sample, and how these combinations effected recipient response behaviour and interactive outcomes. Specifically, I looked at whether information pertaining to the recipient (i.e., visual orientation, kinship to signaller) influenced the production of multimodal combinations by focal individuals and whether multimodal combinations were more effective in soliciting responses and successful communicative outcomes. The results suggest that older individuals have an increasing understanding of recipient's visual orientation and produce multimodal combinations most often when a recipient was visually available. In addition, I found that regardless of signaller age, multimodal combinations were more likely to be responded to, responded to faster, and satisfy apparent communicative goals. Together, in support of the hypothesis that multimodal combinations disambiguate communicative messages, these findings indicate that similar to humans, even young chimpanzees have the capacity to use multimodal combinations to enhance communicative effectiveness.

Consistent with my predictions, the first set of results showed significant effects of age and recipients' visual orientation on the type of communicative signal produced by immature chimpanzees when initiating or re-initiating a communicative interaction with a conspecific. The finding that older individuals had more initiatory signal events

containing multimodal signal combinations aligns with previous findings (Chapter 2) in which a general trend of increasing multimodality was reported across and within communicative interactions irrespective of whether a focal or recipient was the initiator. Evidence that signal events containing multimodal combinations were more likely to be produced when the recipient was visually oriented towards the signaller is consistent with a 'clarifying' function of these signals for which all components can hold complementary information and reinforce a given message to a specific recipient (Taglialatela et al., 2015). For this information to be communicated and for associated effective and successful outcomes to be achieved, the recipient must be able to visually access the message in its entirety. Even vocal components may contain additional visual information only available when visually directed (see Taylor et al., 2022 for more on vocal directedness in chimpanzees). This may help explain our finding that all multimodal combination types were produced more often in the visual field of a recipient than not, irrespective of modality. The apparent proclivity of young chimpanzees to employ multimodal combinations to visually available recipients could therefore be indicative of an ability of young apes to understand that multimodal signals are most effective when they are within a recipient's visual space. In turn, this would suggest that they are capable of intentionally pursuing a communicative goal, as is the case in human infants (e.g., Bates et al., 1975, 1979; Liszkowski et al., 2008; Igualada et al., 2015). Furthermore, the lack of expected interaction found between age and sensitivity to recipient's visual orientation implies that that such an ability is present even early in chimpanzee ontogeny.

Across the second and third set of results, I found a that signal events containing multimodal combinations were more likely to be responded to (i.e., effective) and that

these responses were also more likely to match the apparent communicative goal of the signaller (i.e., successful) at all ages. Previously, the effectiveness of multimodal (vocal-gestural) combinations has been previously explored in *Pan* in terms of frequency or speed of recipients' responses (Genty et al. 2014, 2015b; Wilke et al. 2017). In chimpanzees, vocal– gestural (gesture + 'grunt') combinations are more likely to elicit responses in recipients than vocalisations alone are, but gestures alone are just as likely to elicit responses as combinations are (Wilke et al., 2017). Furthermore, signallers are more likely to switch from single signals to combinations to get a response, only following the failure of a vocal but not a gestural signal (Hobaiter et al. 2017). While these studies suggest that when compared to the gestural component alone, multimodal combinations elicit lower response rates in chimpanzees, the current study differs in that I also considered facial expressions. Moreover, my comparison between signal events containing multimodal vs only unimodal behaviours, included vocal and facial unimodal signals not just gestures. My results suggest that when considered holistically, multimodal combinations of facial expressions, gestures and vocalizations are more effective in eliciting responses in immature chimpanzees than any of these used in isolation more generally. That said, I did not separate signals dependent on modality so it could be that unimodal-gestural signal events were particularly effective, but this effect was lost when grouped together with other unimodal signals. To fully test this, it will be necessary to expand on the current and aforementioned studies to specifically test the effectiveness of unimodal in relation to multimodal combinations, including the role of facial expressions alone and as part of combinations.

Previously, the production of multimodal signal combinations in *Pan* have been linked to specific social functions (Genty et al., 2014, 2015; Hobatier et al., 2017; Wilke et

al., 2017; Oña et al., 2019) (i.e., how it serves the signaller's social goal, identified by recipient's responses across contexts). More recently, Genty (2019) went a step further and identified functional specificity of distinct combination types *within* social contexts, wherein different combinations were used to solicit a specific response in bonobo infants including food-begging, nursing, carry/contact, and reassurance. Additionally, in a previous study (Chapter 3) I found that multimodal combinations were highly contextually specific, but multiple combinations were specific to a given context e.g., play. The result of the current study that multimodal combinations promote the successful achievement of communicative goals is therefore consistent with the suggestion that multimodal combinations are functionally specific signals, but further suggests that this functional specificity aids communicative success. That is, if a combination signal has a clear meaning, the recipient is more likely to interpret it successfully and respond accordingly. In line with Genty's (2019) finding that combinations were functionally specific in infant bonobos (age range 9-42 months), these signals were found to be successful across our sample age range showing that even very young chimpanzees can combine signal elements with intention.

The finding that response latency to signal events containing multimodal combinations decreases with age, however, does imply that even though these signals are more effective regardless of age, there is perhaps a degree of refinement of communicative skill with increasing age, related to signal efficiency. Response latency was coded from the beginning of a signal event, so one possibility could be that event duration was shorter in older individuals, leading to the observed decrease of response latency. However, at present this is speculation as signal event duration was not accounted for in my analysis.

158

In the human literature, the duration of multimodal signals has been shown to become shorter across development as children become more efficient in their communication, an important developmental process to prepare for the tight, time constraints of adult turn-taking conversation (Stivers et al., 2009; Levinson, 2016; Holler et al., 2018). For example, in a recent study Murillo et al., (2018) investigated changes in temporal synchrony between gesture and speech in the developmental period between babbling and two-word productions in human infants to test the hypothesis that multimodal communicative behaviours become more efficient with age. They found that, particularly after the age of 9-months, multimodal communicative behaviours became shorter in duration and became increasingly synchronous with increased overlap of constitute elements. In line with studies showing that response times to human speech accompanied by gestures are faster to speech which is not (Holler et al., 2018; ter Bekke et al., 2020), reduced response latency to multimodal combinations in older individuals here could be evidence of the development of increasing communicative efficiency. If the nature of ape multimodal combinations is communicative clarity at all ages, it could be predicted that a developmental increase in efficiency of multimodal signal production would result in faster response times. Future work should incorporate signal duration into models testing the effect of combination signals on recipient responses times to test this hypothesis and illuminate potential parallels between ape and human development in communicative efficiency.

I originally predicted that multimodal signal combinations would be more likely in interactions with non-kin due to a lower familiarity and increased unpredictability with unrelated recipients. For example, Fröhlich et al. (2020) found that 'multi-articulatory' signals (i.e., >1 body part involved in the production of a communicative act e.g., limbs,

gaze, face, voice) were much less likely between orangutan mother-infant dyads than other interaction dyads, due to the high degree of social tolerance between mothers and offspring. This finding is in line with studies of chimpanzees (Fröhlich et al., 2017) and bonobos (Genty et al., 2015a), which have demonstrated ape signallers are capable of flexibly adjusting their signalling to specific recipients. However, in contrast to this prediction, I did not find that kin relationship influenced the probability of a signal event containing a multimodal combination. One possible factor contributing to this violation of expectation could be the high number of playful interactions observed within our interactions, in which there was a higher likelihood of a multimodal combination being produced. In particular, the 'open mouth' (play) face is an important social regulator in playful interactions (Waller & Dunbar, 2005) and was frequently paired with gestures and laughter vocalizations (Chapter 3) in immature interactions irrespective of kinship.

It is also possible that this null finding was due to limitations in the current study, which was unable to account for potential variation related to affiliative relationships. I recognize that the current analysis would benefit from more nuanced representations of individual relationships and hierarchical positioning when considering the identity of the recipient in relation to the signaller. By only including a broad measure of kin relationship here, I could not account for variation in familiarity, social closeness or rank distance and any associated variation in social tolerance across recipients. All of these may be influencing signalling behaviour. Unfortunately, this information was not available at the time of analysis but moving forward, I believe it important to include data from social network and dominance rank analyses to gain a more accurate representation of relationships before drawing conclusions about the role receivers play in shaping signal use across development.

While this research contributes new insights into the development of multimodality in chimpanzee communication, it is important to acknowledge the potential impact of research setting in the interpretation of these findings. In the sanctuary environment, most observations took place near an enclosure fence line which are typically free of dense vegetation as to prevent chimpanzee escapes. Although this facilitated observation conditions, it also increased visibility between potential communicative partners in this area which may have contributed to more signal events (including multimodal combinations) being produced while a recipient was visually oriented towards a signaller without obstruction. Nonetheless, even in close distance interactions there is potential for a signaller to be out of view, simply by where how are positioned relative to their intended receiver.

So, it remains indicated that multimodal combinations are production with the intention that they be seen in their entirety.

Lastly, most opportunities for observation occurred during pre-feeding periods when the chimpanzees would gather near the fence line. Pre-feeding periods are associated with high social tension, which can lead to an increase in tension mitigating behaviours such as social play (Palagi et al., 2004), particularly within our immature sample. As mentioned, I found that initiative signal events containing multimodal combinations were produced most often within playful interactions. An additional consideration is that along with the signaller, prospective partners were also present near to the cleared enclosure fence line and therefore, recipients were once again more likely to be close by and more able to see signals directed to them. In addition, the play context had a positive effect on multimodal combination effectiveness which once again could be due to a high number of observations in this context and the related visibility between

161

focal and recipient. I acknowledge that the increased opportunity to observe play-interactions in sanctuary pre-feeding periods could be over-estimating our observed effects, which may not be as strong across social contexts in more natural settings.

Conclusions

In conclusion, my findings indicate that immature chimpanzees might have an ability to produce multimodal signal combinations to enhance effectiveness and success of communication. Furthermore, this ability appears to be present even in early ontogeny similar to observed patterns in human infants (e.g., Bates et al., 1979; Liszkowski et al., 2008; Igualada et al., 2015). I found an age-related decrease in response latency when signal events contained multimodal combinations which could reflect an increase in signalling efficiency, but this requires further exploration. Future studies of communicative development in a more natural setting, and a closer examination of the role of distinct combination types will add clarity on the degree of intentionality and flexibility young apes have in their communicative productions and the role of multimodal combinations in shaping outcomes and signalling behaviour across time.

Chapter 6 - General Discussion

In this chapter, I reflect upon and discuss my empirical findings from previous chapters and their contribution to our wider understanding of multimodal communication development in chimpanzees. In each section, I also consider the limitations of my work and how future studies could address these shortcomings, as well as new avenues of research which would advance our understanding of this topic.

Overview

In the opening chapter for this thesis, I highlighted the importance of comparative-developmental perspectives in communication research to understand the selective pressures that shape the evolution of communication systems. The evolution of observable behavioural phenotypes occurs via developmental processes that affect phenotypic endpoints on the ontogenetic level (Müller, 2007). Therefore, investigating the development of multimodal communication in our closest living relatives can shed light on the evolutionary origins of the multimodality that characterizes human language. So far, the adoption of multimodal approaches to great ape communication research is still relatively new and as such, existing research is still limited. While emerging research has been fruitful in exposing the occurrence and apparent function of these signals within the repertoires of adults (e.g., Pollick et al., 2008; Leavens et al., 2010; Genty et al., 2014; Taglialatela et al., 2015; Wilke et al., 2017) to date no other study has yet focused on the development of these signals across young ape ontogeny. Thus, with this thesis my primary goal was to address this gap, taking a comparative-developmental approach to examine multimodal communication in a cross-section of immature chimpanzees.

As demonstrated in Chapter 1, the literature surrounding communicative development of chimpanzees, as well as communication in general, is generally characterized by a focus on gestural, vocal or to a lesser extent facial signals in isolation. This divide has also led to individualized perspectives on the proximate and ultimate mechanisms which underpin the acquisition and development of signal repertoires at various stages of ontogeny. While facial expressions and vocalizations have traditionally been considered innate and emotionally driven (Liebal & Oña, 2018), conflicting views on the role of genetic and social factors in shaping the development of gestures are found across research groups and associated approaches (Liebal et al., 2019). While apparently different findings may indicate that more than one mechanism is involved in the development of gestural communication, emerging research in the facial (e.g., Davila-Ross et al., 2015; Lembeck, 2015; Waller et al., 2015) and vocal domain (e.g., Wilson et al., 2007; Slocombe & Zuberbühler, 2007; Crockford et al., 2012; Watson et al., 2015; Dezecache et al., 2021) adds incentive to consider that these signals are also subject to an element of control and as such may be subject to learning. Furthermore, as the proclivity to combine these signals into singular messages is now well documented in chimpanzees, it is equally important to consider the role of multimodal communication. Uncovering the developmental pathway of multimodal signal combinations allows us to compare patterns of acquisition and usage of communicative modalities when produced separately vs together, which may have independent trajectories. Together, this approach will lead to a better understanding of the complexity or integrated nature of developmental processes shaping primate communication systems.

To this end, in this thesis I examined patterns of communication development in semi-wild, immature chimpanzees, focusing on the production of multimodal signal

combinations. Using a cross-section of individuals ranging in age from early infancy to the brink of adulthood, I investigated whether age contributed to patterns of signal use (Chapter 2) while also considering the influence of socio-ecological variables including context (Chapter 2) and interaction behaviour with mothers (Chapter 3) and the wider social community (Chapter 4). Lastly, I examined the role of multimodal signal combinations within the interactions of young chimpanzees (Chapter 5) to shed light on the potential roximate mechanisms underlying these signals at different stages of development.

Developmental trajectory of multimodal signal combinations

In my first empirical chapter (Chapter 2), I examined the development of multimodal signal combination production in relation to unimodal signals using a cross-sectional sample of immature sanctuary-living chimpanzees. I found that older individuals used multimodal combinations significantly more frequently, and at higher relative proportions than younger individuals, while unimodal signals were produced at consistent frequencies across all ages. In addition, I recorded a comprehensive repertoire of 101 multimodal combination types (Chapter 2), which appear strongly linked to behavioural context (Chapter 4). Nonetheless, unimodal signals remained the dominant form of communication regardless of age. This finding mirrors previous observations in wild chimpanzees (Wilke et al., 2017), where multimodal signal combinations were also found to be produced at consistently lower rates than unimodal signals in individuals aged from 8-40 years old.

Elements of these findings both reflect and contrast with patterns seen in human (illocutionary) communication development. Where they are comparable, is that across development, human infants exhibit increasing articulated control over the

combination of intentional gestures and speech to convey more specific

communicative meaning (e.g., Iverson & Goldin-Meadow, 2005; Esteve-Gibert &

Prieto, 2014; Murillo et al., 2018). As such, there is an increasing pattern of

simultaneous gesture-speech coordination, which eventuates into the deeply integrated

gesture-speech system of human adults (Kendon, 1980; McNeill, 1992; De Ruiter,

2000; Kita, 2000). That said, where my findings differ with patterns of human

development is that unimodal signalling remains the most frequent form of

communication in chimpanzees across all age groups I sampled, rather than these

signals being replaced by a predominantly integrated communication system. This

apparent discrepancy in developmental patterns may be representative of species

differences in sociality for which different affiliative strategies are required to

maintain social relationships in social systems with varying degrees of complexity.

Communicative complexity has previously been hypothesised to be a consequence

of the high cognitive demands of living in complex societal structures (Freeberg et al.,

2012). The primary mechanism through which all group-living primates maintain social

bonds is via mutual grooming (Dunbar, 1991, 1996). However, as group sizes increase, the

time and energy budgets required for grooming become increasingly demanding and

inefficient given higher foraging competition and predation risks. As such, with increased

group size, group cohesion can decay leading to group-splitting (Henzi et al., 1997).

Chimpanzees live in fission-fusion societies with complex social dynamics. In these

multilevel societies, close social bonds are also maintained via reciprocal relationships

marked by increased energy investments in repeated and reciprocated instances of

association and interaction e.g., grooming (Watts, 2006; Mitani, 2009). However, recent

research has provided insight into how chimpanzees manage the high energetic demands

associated with maintaining these group relationships. For example, Roberts and Roberts (2016a, b) found that in addition to higher rates of mutual grooming, individuals who spend longer time in close proximity also have higher rates of reciprocal, affiliative gestural communication. The authors suggest that these affiliative signals may function as a form of tension mitigation associated with longer periods in proximity but also be a form of "grooming at a distance". In this respect they may exploit on a larger scale the psychopharmacological mechanisms involved in grooming behaviour (Dunbar, 1996; Tarr et al., 2016) while simultaneously mitigating the energetic costs. Furthermore, Roberts and Roberts (2019) also found that a greater number of close proximity bonds within larger social networks was associated with more complex communication (i.e., gestures accompanied by self-relevance cues and gesture-vocal combinations) further suggesting that communicative complexity is a means through which to maintain bonds with a higher number of conspecifics where grooming would be wholly inefficient.

During hominin evolution, the increased flexibility in the use of different types of signal combinations may therefore have facilitated the maintenance of different types of social relationships and social cohesion in large multi-level groups, acting as an alternative to other bonding mechanisms that require physical contact and higher energy budgets such as grooming (Dunbar, 2012; Freeberg et al., 2012). While the basic properties of multimodality are present in chimpanzees, significant expansion of neural structures (e.g., the neocortex and cerebellum) and associated cognitive capacities paired with increasingly complex and varied communication networks may have helped larger groups of humans maintain social cohesion and coordinate their activities serving as a more efficient bonding mechanism to grooming (Dunbar, 1993).

For this thesis, I did not have access to long-term data regarding proximity bonds and social networks of individuals. However, tying this kind of information together with patterns of combination signalling across ontogeny would be a fascinating avenue of further research. Future work would benefit from exploring the links between sociality and network size in chimpanzees and the rates of multimodal signal combinations and how different signal types are used to regulate different types of social relationships. In addition, it would be interesting to compare rates of multimodal signalling in wild chimpanzees living under- natural conditions in groups of varying sizes. This would allow us to better understand the role of social complexity in providing a space for complex language to evolve, thus providing a valuable 'bottom-up' perspective that could complement and inform our interpretation of comparative studies.

While differences in *Pan-Homo* social complexity may help explain the persistent dominance of unimodal signals in chimpanzee communication, my findings overall reflect a developmental shift in multimodal communication. That is, individuals as young as 1 year old possess the capacity to produce multimodal signal combinations, but that older individuals do so significantly more frequently and at higher relative probabilities (Chapter 2). However, to a degree, changes in social complexity or rather, changes in the complexity of individual social networks and exposure to new interactional contexts with increasing age, could also help explain increases in combination signalling. For example, given the evidenced 'disambiguation' function of great ape multimodal combinations (e.g., Genty et al., 2014, 2015b; Hobaiter et al., 2017; Wilke et al., 2017; Oña et al., 2019), these signals may become increasingly required in interactions with less predictable outcomes or in contexts in which the cost

of miscommunication is particularly high e.g., physical aggression. In this sense, both

biological and comparative psychological perspectives highlight the value in integrating

bottom-up' and 'top-down' approaches to animal communication to get a fuller picture

of the selective processes.

Clarifying the role of context

A running theme throughout each empirical chapter of this thesis is the influence of

behavioural context on multimodal communication, particularly the contexts of play and

aggression. Both contexts appear to promote more frequent production (Chapter 2) and

more variation in combination types (Chapter 4). Multimodal combinations are also

produced at higher relative probabilities than unimodal signals during aggressive contexts,

(Chapter 2) and are responded to more frequently in play (Chapter 5). In addition, my

results showed that chimpanzee mothers appeared more motivated to attend to infant

signals in more 'urgent' contexts which included when infants were distressed or

responding to aggression within the group (Chapter 3). As I touched on in these chapters,

whilst these findings highlight the relevance of these behavioural contexts in promoting

signal production and specifically combinations, broad-category contextual classifications

likely do not capture the subtle nuances of an interactional situation.

Indeed, one of the key issues that still needs to be addressed in future studies

concerns the clarification of 'context' in relation to the production of communicative

signals. As it stands, most comparative work on communication has used broad

behavioural categories when considering context. However, there are many more

immediate factors which could contribute to signal choice and interaction outcomes. For

example, with the aim of consistency, my definition of an aggressive context was based

upon previous descriptions in the ape literature (Schneider et al., 2012; Hobaiter et al., 2017) and included 'initiation of or responses to aggressive behaviours including threats or physical agonistic encounters'. I also included here submissive signals as these also function to mitigate the threat of aggression from dominant individuals (Smith, 1977). It is important to recognize that different elements of this definition may themselves be sub-contexts in which an individual may be more or less likely to produce a certain signal such as a multimodal combination. Furthermore, signal production could also vary depending on the cause of the aggressive interaction e.g., food-related vs submissive greeting.

This is a complicated issue as in addition to the above, context may vary according to whether it is the perspective of the initiator, the recipient or the general setting that is being considered. One possible way to help capture more accurate contextual information in relation to communication could be to focus more on the *communicative* context, by focussing more on the communicative goal of the signaller. In gestural research, the method of 'goal-outcome' matching (Cartmill & Byrne, 2010) and assessing 'apparently satisfactory outcomes (ASO's)' (Hobaiter & Byrne, 2014; Graham et al., 2018) has been used to identify the meaning of certain gestures by tracking the most common recipient responses/ interaction outcomes that satisfy the signaller as deduced by the cessation of signalling. While I used the goal-outcome matching method to inspect the role of multimodal combinations within communicative interactions more generally (Chapter 5), I did not have enough consistent repetitions of combination *types* to test their individual effect on recipient responses and interaction outcomes. However, in the future, incorporating this method alongside the coding of wider communicative context could help break down the specific types of interactions within wider contexts in *which* multimodal

combination types are most frequently produced, and where they appear to carry a particular communicative function. In addition, as well as coding behavioural contexts (e.g., aggression) and communicative meaning (e.g., stop that) controlling for the interaction initiator, the familiarity and rank distance between interactants could provide a much clearer picture of the circumstances which encourage the production of more complex, multimodal signals and how this varies at different stages of ontogeny. Unfortunately, I did not have access to such data for inclusion in my analyses but acknowledge the importance of social-affiliation and hierarchical information in forming more accurate interpretations of behavioural patterns in the future.

Lastly, I highlighted in Chapter 2 that the contexts in which I most observed multimodal signal combinations (play and aggression) contrasted with research on wild chimpanzees (Wilke et al., 2017), in which the contexts in which they were recorded in sufficient numbers to be analysed included rest, groom and feeding. As I discussed previously, this is potentially due to differences in observation conditions between the wild and (semi-wild) captive research settings. Currently, the current lack of systematic multimodal research makes it hard to make deductions about which of these contexts influence the production of combination signals most or whether they may vary according to socio-environmental conditions. That is, we don't yet know if the respective findings from each study represent behavioural differences across research settings or are more a product of observation conditions obscuring the recording of communication within certain contexts at each site. Expanding the search for variation in multimodal communication across multiple socio-ecological environments, including but not limited to the wild, will be crucial to address outstanding questions surrounding the role of context in multimodal signal combination production across ontogeny, as

well as its function in general.

Proximate mechanisms underlying patterns of multimodal combination signalling

While results from my first empirical study (Chapter 2) showed that multimodal combination frequency increased significantly with age, in Chapter 4 I did not find that older individuals with higher mean communicative interaction rates used more signal combinations within these interactions. Initially, I expected the opposite pattern predicting that individuals who engaged in communicative interactions more frequently would have more opportunity to practice complx signals, therefore promoting their production. This was the case in the recent gestural study by Fröhlich and colleagues in 2017, who found interaction rates positively predicted increases in the number of gestures produced within communicative interactions. The lack of an effect of communicative interaction rates on the production of multimodal combinations may be the result of study limitations or indicate that different mechanisms underlie the production of these signals. Specifically, it seems these signals are more related to social context, having an important function in message clarification regardless of signaller age.

It could be the case that production of different signal components e.g., gestures, are positively affected by interactive behaviour across ontogeny but their use in multimodal combinations are more related to circumstances of production. For example, the inclusion of a high number of potentially more arousal-based signals like facial expressions (e.g., play faces) or vocalizations (e.g., squeaks or grunts during submissive interactions) in multimodal combinations (Chapter 2) could have contributed to a more even spread of combination frequency in my sample, irrespective

of interaction rates. Instead, I found that the number of conspecifics with whom focal individuals interacted with had more of an influence on combination production. This too could be capturing more interactional context in terms of who is the recipient. As discussed, variation in relationships, rank distance or age could affect the production of multimodal combinations. One could be interacting at high rates with certain individuals but depending on who they are and the context of the interaction, unimodal signals may be sufficient to communicate a message; this could enhance the effects of mean interaction rates previously observed in the production of gestural signals alone (Fröhlich et al., 2017). However, different relationships and contexts may need to be navigated with more caution contributing to the strong effect of these variables on combination production.

Although not touched on in this thesis, examining the developmental trajectory of distinct combination types could help disentangle the influences of these different interactional effects. It could be that combinations containing gestures are affected differently than facial-vocal combinations in terms of production across time. Examining the types of gestures used within multimodal combinations at different stages of ontogeny could reveal a pattern more akin to previous unimodal work. Furthermore, incorporation of social network matrices into investigations of communicative signalling would help us understand the role of recipient, in conjunction with wider behavioural context, in signalling tactics.

My finding that combination types appear highly contextually specific is an important contribution to our understanding of communicative flexibility in apes. Previous gestural studies (Roberts et al., 2012; Hobaiter & Byrne, 2011a, b, 2014; Byrne et al., 2017) describe that chimpanzee gestures are produced flexibly across

contexts with authors theorizing that meaning is subsequently extracted from the context in which they are produced. But what they do not consider is *additional* signal modalities that may also be present in combination with a given gesture. The context specificity that I found in multimodal combinations (Chapter 4) implied that if gestural research did consider other modalities, perhaps the flexibility of the 'holistic' signal would not be as high as when the gestures are observed alone. This is a critical point for future gesture research. In this respect, my findings in Chapter 4 do not necessarily refute the evidence from gestural studies alone i.e., the gestural components may indeed be flexible, but it highlights the importance of multimodal research in the interpretation of behavioural patterns.

High contextual specificity of multimodal combinations may indicate that different combination types hold specific message meaning, which fits with the finding that combinations were more likely to be responded to and achieve communicative goals (Chapter 5). The idea that multimodal combinations become more important with increased exposure to more functionally critical contexts would fit with the general increase in signal combination frequency described in Chapter 2, but I did not find an interaction between these variables. As I discussed in Chapter 2, this could be due to the high number of playful interactions in my sample in which multimodal combinations were prevalent so it would be interesting to see whether future studies and expansion of datasets would in fact reveal such a trend.

It is important to note that across my empirical chapters, I repeatedly found no significant interaction terms between age and my key model predictors including context (Chapter 2); mean interaction rates and number of interaction partners (Chapter

4) on multimodal combination frequency (Chapter 2, 4) and repertoire size (Chapter 4); and multimodal combinations in relation to predicting goal-match outcomes (Chapter 5). Instead, these variables were all significant regardless of age. This makes it more difficult to interpret developmental mechanisms and highlights an important limitation of my cross- sectional sample. It would be beneficial to examine patterns of interaction behaviour over longer periods prior to the observation of communicative behaviour to account for variation in these patterns which may not be represented in shorter observation periods. Unfortunately, with the effect of the pandemic on international travel and a restricted ability to conducted long-term observations, I was unable to include such data at this time, however I acknowledge the clarity continued data collection could contribute to my findings.

Lastly and importantly, here I would like to acknowledge that some of the models included in this thesis were not subjected to explicit checks for statistical power, and while I attempted to control for over-parametrization via various diagnostics this is still a consideration in the interpretation of my more complex models. It is possible that these factors have influenced the lack of age and context-related findings across various chapters. This highlights the need for caution, and I recognize the necessity of more rigorous exploration in future research endeavours.

Further avenues for future research

While the empirical studies in this thesis provide novel data and insight concerning the production of multimodal signal combinations across chimpanzee ontogeny, they also come with some limitations. As multimodal research is still in its infancy, this body of work is only a starting point: there are many directions in which to develop our

understanding of complex communication in chimpanzees and other ape species, as well as to build comparability with the fields of human psychology and linguistics. I have already discussed key limitations and some suggestions for future lines of study in previous chapters, therefore rather than repeat, I will expand on these with additional directions here.

Quantifying structural variation within signals

In human language, verbal utterances identical in syntactic structure and semantic content can be interpreted completely differently depending on the accompanying acoustic and visual information. For example, an utterance such as 'they are coming' could reflect the speaker's excitement, surprise or irritation of this fact depending on the tone and pitch of their voice, but also visual signals such as facial articulation (e.g., smiling vs furrowed brow), and body language (e.g., standing upright vs bowing of the head). That is, there is an interactive relationship between audio and visual components of a single utterance that contributes to its meaning.

In this thesis, I relied on established ethograms of communicative behaviours across multiple signal modalities. As it stands, previous multimodal research has thus far relied on the examination of broad-category signal combinations e.g., gestures, call types and prototypical facial expressions, identifiable at the observation level. However, recent research in the vocal domain has demonstrated that during vocal development, young chimpanzees show increasing acoustic variability within call types (Taylor et al., 2021) and distinct acoustic variants have been linked to different functions (Dezecache et al., 2021). Furthermore, both acoustic variants *as well as* the addition of visual signals (i.e., gestures) to specific acoustic variants have been shown to modulate the meaning

of chimpanzee calls (Fedurek et al., 2021). In addition, while chimpanzee facial expressions have been categorised as prototypical displays (Parr et al., 2007), it is also acknowledged that they can be highly blended and graded between prototypical descriptions (Parr et al., 2005; Burrows et al., 2006). Lastly, the requirement for gestures to meet certain intentionality criteria (Hobaiter & Byrne, 2011a) has led to a potential over-looking of bodily movements or postures which may be more involuntary but nonetheless add communicative information (see next section).

An example of just how flexible a 'single' signal can be has been demonstrated via the 'touch' gesture commonly produced by chimpanzees (Bard et al., 2019). The authors demonstrated that this gesture type alone had 36 different forms, directed across 70 different target locations of a social partner's body and was produced in 26 different interactional contexts. Also, infants (2–5 years old) differed significantly from the adults in form, location, and context of their touch gestures indicating ontogenetic shifts in usage. It was not found that there was a high degree of contextual specificity of form-target location patterns within different contexts which aligns with previous studies of chimpanzee gestures citing their contextual flexibility (Hobaiter & Byrne, 2014; Byrne et al., 2017). However, the presence of additional visual (e.g., facial expression, body postures) or vocal signals were not considered so it could be the case that information related to contextual specificity (e.g., Chapter 4) or message meaning were missed. Nonetheless, this study highlighted the protentional form-flexibility of a single signal which many studies (including in this thesis) would not consider using ethograms alone. Expanding studies like Bard et al.'s, (2019), considering form, target, individual and contextual information accompanying distinct signal variants (including combinations) will help identify message meaning and modulating effects of different added elements.

This will be a great feat and require considerable effort but is a truly exciting avenue which could reveal so much about the communicative world of our closest relatives.

Given that multimodal perception in chimpanzees has demonstrated that like humans, they can recognize audio-visual correspondences between their various call types (Izumi & Kojima, 2004; Parr, 2004), we know that distinct message streams also hold important communicative value for the interpretation of all available information. Paired with the potential for high levels of form flexibility within different signal types, it is evident that we have barely begun to understand the full extent of structural complexity of multimodal signals in our ape relatives. If distinct acoustic/ visual variants of signals interact in meaning-modulation when paired together, just imagine what we are missing beyond what is visible through observational study alone. As this topic progresses collaborative efforts to quantify as much audio/visual structural variation in different signals as possible would help in identifying the extent to which elements can overlap.

In one example of how such quantification could be approached, Smith and Evans (2013) offer a heuristic for identifying and interpreting how signals across various sensory modalities interact to determine signal function and efficacy. Their approach explores how combinations of sensory modalities can either be redundant (i.e., containing the same information when produced separately or together) or introduce nonredundant information content, ultimately influencing the function of the combined signal. The authors propose a geometric framework for visualizing receiver responses to multimodal signals in a three-dimensional space via surface plots, facilitating the understanding of complex displays. For instance, this framework can be applied to study how auditory and visual components in a male courtship display impact female preference when presented separately or at different

levels of combination. While Smith and Evans primarily discuss 'multimodality' from the perspective of sensory perception, their heuristic could also be adapted to examine functional variation through the combination of distinct communicative behaviours, such as gestures, facial expressions, or vocalizations. However, it becomes more challenging when dealing with so-termed 'multi-component multimodal signals' that combine sensory (e.g., auditory, visual) and behavioural (e.g., gestural, vocal) elements. In their definition, 'multicomponent' signals are akin to multimodal signal combinations in this thesis, where distinct communicative signals or behaviours co-occur and which can also occur within the same sensory channel (e.g., visual gesture + facial expression). Exploring scenarios with multiple sensory and behavioural components simultaneously introduces complex geometric mapping and multi-level effects on recipient behaviour and introduce a whole new understanding of the intricate, multimodal structures within animal communication.

The question of intention

Communicative intentionality is a key feature of language (Tomasello, 2008). Instead of producing sentences as automatic responses to stimuli, their production represents an intention to alter the behaviour or mental state of other individuals (Grice, 1969). During development, human infants increasingly communicate in goal-oriented manner (i.e., *first- order* intentionality) and this is associated with the ability to understand interactive partners as social agents (Bates et al., 1975) with independent thoughts and beliefs (i.e., *second- order* intentionality). In non-human primates, distinct behavioural criteria based upon research of pre-linguistic communication in human infants have been established to investigate communicative intentionality and distinguish intentional signals from unintentional acts. In particular, three behavioural markers have been proposed to be reliable estimators of communicative intent: audience checking,

persistence to the goal and sensitivity to recipient's attentional state (Leavens et al., 2005; Townsend et al., 2017).

In great ape gestural research, there has been a reliance on these markers of intentionality to discriminate between communicative vs noncommunicative bodily movements. In fact, across studies the definition of a gesture is rooted in the understanding that it is an intentionally communicative act. This contrasts with facial expressions and vocalizations which are more often considered emotionally driven, although have also been evidenced to be under at least some degree of volitional and intentional control (facial expressions: Lembeck, 2015; Davila-Ross et al., 2015, vocalizations: e.g., Crockford et al., 2012; Schel et al., 2013; Watson et al., 2015 Dezecache et al., 2021; Leroux et al., 2021). That is, an 'open mouth' (play) face could be the arousal-based product of excitement during a playful interaction but could equally be used as part of an intentional signal (e.g., with a slap other gesture) to initiate a playful interaction in the first place. While in the latter case it would also be accompanied with intentional markers associated with the gesture, in both cases the facial expression would have recognized communicative value and coded in the same way. This would not currently be so for a gesture.

This assumption that gestures are, by definition, intentional but not other signals is problematic if we are interested in understanding the cognitive mechanisms underlying multimodal signal combinations in ape species, and how these are shaped across ontogeny. To do so, we should be able to discriminate between signals which are learned/ intentionally produced and those which are innate/ bound to affective states. As it stands, I am aware that my dataset likely captures a spectrum including combinations

where (1) all components are intentional, (2) part of the combination is intentional, and part is arousal-based and (3) all components are arousal-based. In Chapter 5, although my findings may indicate a degree of intentional control of multimodal signal production, I have been careful with my language and make no claims at present that this is a fact, particularly across individual components. It is a difficult issue to address and is part of the complicated nature of multimodal research. In addition, if facial expressions and vocalizations do not have to be intentional to be communicative, can the same be said for gestures? Once again, ape gestures have been synonymous with intentionality as their very identification now typically relies on at least one accompanying marker of intentional production. However, in studies of human communication and emotion recognition, the communicative value of a gesture is not contingent on it being accompanied by these specific intentional markers. Even when facial and vocal information is excluded, humans are able to identify communicative sentiment via skeletal movements and posturing, for example hands being raised to the head or mouth during surprise, clenched fisted during anger or a bowed head and dropped shoulders when apologetic (Noroozi, et al., 2018).

Regardless of the intentionality, such signals could still have an important social function (e.g., disambiguation) and studying how different socioecological pressures influence their production across development could still have implications for our understanding of the evolution of these signals before and after the *Pan-Homo* split. In perlocutionary acts, communication functions via a receiver interpreting the behaviour of the 'sender' while in contrast, illocutionary acts are produced with the aim of carrying out a socially recognized function ('intentional signals'). While a question mark remains over how to distinguish intentional vs arousal-based facial and vocal

signals, one possible way to quantify multimodal signals with perhaps varying degrees of intentionality across communicative components moving forward could be to systematically integrate the 'production' vs 'perception' perspectives in multimodal research.

An example of how this could be done can be found in a recent study by Fröhlich et al. (2021). The authors studied the communicative signals produced within wild Orangutan dyads with varying degrees of familiarity and created mutually exclusive communicative categories involving different communicative acts (components) (i.e., socially directed, mechanically ineffective movements of the face, body or vocalizations) and sensory channels (i.e., sensory modality through which a communicative act is perceived: visual, tactile, auditory, seismic). This definition of a communicative act thus allows for both intentional and unintentional communicative information contained within a given signal to be accounted for and the contribution of constituting elements to be assessed. The resulting categories included (1) uni-component unisensory acts, (2) multicomponent unisensory acts, (3) uni-component multisensory acts and (4) multicomponent multisensory acts. Using this coding scheme, the authors were able to investigate differences in production and outcomes of constituent signal parts and found that the integration of different communicative acts functions to disambiguate a message (i.e., specify a meaning), whereas the integration of different sensory modalities serves to ensure the message arrives (i.e., enhance effectiveness).

Again, the multimodal combinations coded within my dataset for Chapters 2, 4 and 5 likely contain various combinations of intentional/ unintentional components,

which may be subject to different ontogenetic pressures and add independent value to the holistic signal. In Chapter 5 I found that multimodal combinations, as they are defined in this thesis, both solicit more responses and goal-match outcomes. If both communicative acts and sensory modalities were considered in my analyses, perhaps this finding could have been split between those which contained multiple sensory modalities vs multiple components which would align with Fröhlich et al.'s (2021) research. Whether that be the case or not, in future research it will be extremely valuable to conduct multimodal research at the interface of perception- and production-focused research and bring together coding methodologies that could be applied across respective research disciplines. Addressing the current gap between production and perception features of communicative acts in this way will enable a greater degree of comparability with human communication and the origins of our multimodal communication system.

The sanctuary environment

As touched on throughout the chapters of this thesis, a final aspect of this research I want to further acknowledge is the influence of the sanctuary setting on my empirical studies and interpretation of their findings. Chimpanzees, like all great apes, are endangered (Humle et al., 2016). Across the last few decades, habitat loss, poaching and disease have led to the decimation of wild populations and the filling of sanctuaries (PASA, 2023) who aim to rescue, rehabilitate and care for their residents. At Chimfunshi Wildlife Orphanage, where data collection took place, chimpanzee groups live in large, naturalistic enclosures allowing for the engagement in naturally occurring social and ecological behaviours. Conditions at these sanctuaries offers certain practical advantages over observations in the wild, which may help researchers capture larger datasets of socio-communicative behaviours, which

themselves are often rare or difficult to see in the wild. For example, the twice-daily occurrence of extra food provisioning (outlined in Chapter 2) and associated proximity individuals maintain to the enclosure fence lines, mean that visibility of social interactions and communicative behaviour is likely much enhanced in comparison to observation conditions in wild settings. This ability to collect large amounts of behavioural data can help expose patterns of behaviour at an accelerated rate, which can then be used to formulate key questions to be addressed in more ecologically valid and evolutionarily relevant environments.

However, the very role of sanctuaries in offering care and refuge to often traumatised young apes means that their subsequent social behavioural development can be altered (e.g., Freeman & Ross, 2014; Llorente et al., 2015; Crailsheim et al., 2015). As such, patterns of behaviours cannot be generalized as species typical. Even though all individuals in my sample were born at the sanctuary, and all but one raised by their biological mothers, I have no way to control for the potential trickle-down effects of matrilineal trauma on infant rearing and behavioural development which could influence study findings. That said, in a recent study van Leeuwen et al. (2022) found that orphan chimpanzees exposed to early social trauma, were socially indistinguishable from mother-born individuals raised at Chimfunshi, in terms of average party size, social proximity networks and grooming frequency. While specific to this location, their findings suggest that sanctuaries can be valuable rehabilitation centres for orphaned chimpanzees, facilitating chimpanzees' potential to cope with early life adversities.

Furthermore, even in sanctuaries such as Chimfunshi which have gone to great efforts to mimic naturalistic environments, the large but confined enclosures restrict residents' ability to engage in fission-fusion subgrouping as they would in the

wild and rely largely on human provisioning to meet nutritional requirements. Food provisioning results in group aggregation around the feeding (and key observation) area and can contribute to heightened group tensions associated with impending feeding competition which likely contributed to the excess of play and aggression related communication I observed in my sample. While the production of multimodal signal combinations in these contexts conforms to previous research citing the role of combinations in signal clarification, their strong effect in my sample may not represent the experiences of individuals developing in the wild. Thus, even though these findings offer insight into social drivers of combinatorial signalling, differential exposure to these contexts in wild groups with fission-fusion dynamics could lead to different patterns of observed communicative development.

Conducting research at sanctuaries can have valuable secondary impacts on study species. The financial contribution of research fees can help accredited sanctuaries fund the care of resident animals, as well as help raise awareness to the threats faced by wild primates, invest in local communities through job provision and education, and tackle illegal activity surrounding the capture and trafficking of different species. However, findings from the empirical chapters presented in this thesis should be considered in the context of these conditions. It will be important to use research conducted at sanctuaries to address resulting hypotheses in wild populations to attain greater ecological-validity in the investigation into natural communication development in chimpanzees.

Conclusion

Studying chimpanzee communicative development from a multimodal perspective has

revealed that chimpanzees increasingly use multimodal signal combinations with age, which is likely linked to age-related changes in the contexts to which they are exposed. Early social input from mothers appears to support a transition from action-based to illocutionary signalling in early infancy which subsequently shaped in interactions with conspecifics that positively influence multimodal signal production and repertoire size. Multimodal combinations are contextually specific and more effective in eliciting responses and achieving apparent communicative goals. Collectively, this work has revealed new evidence that the development of combinatorial communication in chimpanzees resembles some aspects of the pattern seen in humans, while also highlighting important differences between our species. Furthermore, multimodal communication development appears to be influenced by varying socio-environmental factors, and multiple mechanisms may be involved in shaping their manifestation in the chimpanzee communicative system.

In the field of comparative psychology, the evolution of language is typically studied via a top-down approach, searching for shared abilities between humans and our closest primate relatives. Avenues for future research include integrating the different perspectives of multimodal communication and considering how group structure and sociality in natural conditions lay a foundation for the evolution of language from a bottom-up approach. This would further help us understand the emergence of communication in our hominin ancestors and distinguish drivers of variation at the species- or group-level from individual-level which is subject to developmental input.

References

Acredolo, L., & Goodwyn, S. (1988). Symbolic gesturing in normal infants. *Child Development, 59*(2), 450-466.

Adamson, L. B., & Bakeman, R. (1984). Mothers' communicative acts: Changes during infancy. *Infant Behavior and Development, 7*(4), 467-478.

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (2014). Patterns of Attachment: A Psychological Study of the Strange Situation.

Altmann, J. (1974). Observational study of behavior: sampling methods. *Behaviour, 49*(3- 4), 227-266.

Altmann, S. A. (1962). A field study of the sociology of rhesus monkeys*, Macaca mulatta.* Annals of the New York Academy of Sciences.

Appelbaum, M., Batten, D. A., Belsky, J., Boller, K., Friedman, S., Phillips, D., ... & Vandell, D. L. (2000). The relation of child care to cognitive and language development. *Child development*, *71*(4), 960-980.

Arbib, M. A., Gasser, B., & Barrès, V. (2014). Language is handy but is it embodied?. *Neuropsychologia*, *55*, 57-70.

Baayen, R.H. (2008). Analyzing linguistic data. A practical introduction to statistics using R. Cambridge University Press, Cambridge.

Bard, K. (2004). Emotions in chimpanzee infants: The value of a comparative developmental approach to understand the evolutionary bases of emotion. In *Emotional development*. Oxford University Press. Bard, K. A. (1992).

Intentional behavior and intentional communication in young free- ranging orangutans. *Child development*, *63*(5), 1186-1197.

Bard, K. A. (1998). Social-experiential contributions. *Intersubjective communication and emotion in early ontogeny*, (3), 208.

Bard, K. A. (2003). Development of emotional expressions in chimpanzees (*Pan troglodytes*). In *Emotions inside out: 130 years after Darwin's the expression of the emotions in man and animals* (pp. 88-90). New York Academy of Sciences.

Bard, K. A. (2007). Neonatal imitation in chimpanzees (*Pan troglodytes*) tested with two paradigms. *Animal Cognition*, *10*, 233-242.

Bard, K. A., & Leavens, D. A. (2009). Socioemotional factors in the development of joint attention in human and ape infants. In *Learning from Animals?* (pp. 101-116). Psychology Press.

Bard, K. A., & Leavens, D. A. (2014). The importance of development for comparative primatology. *Annual Review of Anthropology, 43*, 183-200.

Bard, K. A., Dunbar, S., Maguire-Herring, V., Veira, Y., Hayes, K. G., & McDonald, K. (2014). Gestures and social-emotional communicative development in chimpanzee infants. *American Journal of Primatology, 76*(1), 14-29.

Bard, K. A., Gaspar, A. D., & Vick, S.-J. (2011). Chimpanzee faces under the magnifying glass: Emerging methods reveal cross-species similarities and individuality.
Personality and temperament in nonhuman primates, 193-231.

Bard, K. A., Maguire-Herring, V., Tomonaga, M., & Matsuzawa, T. (2017). The gesture 'Touch': does meaning-making develop in chimpanzees' use of a very flexible gesture?. *Animal cognition*, *22*, 535-550.

Bard, K. A., Maguire-Herring, V., Tomonaga, M., & Matsuzawa, T. (2019). The gesture 'Touch': does meaning-making develop in chimpanzees' use of a very flexible gesture?. *Animal cognition*, *22*, 535-550.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language, 68*(3), 255-278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv*:1406.5823.

Bates, E. (1975). Peer relations and the acquisition of language. In M. Lewis & L. Rosenblum (Eds.), Friendship and peer relations (pp. 259–292). New York: Wiley.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L. & Volterra, V. (1979). The Emergence of Symbols: Cognition-Communication in Infancy. New York NY: Academic Press

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer quarterly of behavior and development*, *21*(3), 205-226.

Bednarek, M. (2009, November). Corpora and discourse: A three-pronged approach to analyzing linguistic data. In *Selected proceedings of the 2008 HCSNet workshop on designing the Australian National Corpus* (pp. 19-24). Somerville, MA: Cascadilla Proceedings Project.

Beehner, J. C., & Lu, A. (2013). Reproductive suppression in female primates: a review. *Evolutionary Anthropology: Issues, News, and Reviews*, *22*(5), 226-238.

Behringer, V., Deschner, T., Deimel, C., Stevens, J. M., & Hohmann, G. (2014). Age-related changes in urinary testosterone levels suggest differences in puberty onset and divergent life history strategies in bonobos and chimpanzees. *Hormones and behavior*, *66*(3), 525-533.

Bekken, K. E. (1989). Is there" motherese" in gesture? (Doctoral dissertation, The University of Chicago).

Boesch, C. (2009). *The real chimpanzee: sex strategies in the forest*. Cambridge University Press.

Bogin, B. (1990). The evolution of human childhood. *Bioscience*, *40*(1), 16-25.

Bornstein, M. H., & Tamis-LeMonda, C. S. (1989). Maternal responsiveness and cognitive development in children. *New Directions for Child and Adolescent Development, 1989*(43), 49-61.

Bornstein, M. H., Tamis-LeMonda, C. S., Hahn, C. S., & Haynes, O. M. (2008). Maternal responsiveness to young children at three ages: longitudinal analysis of a multidimensional, modular, and specific parenting construct. *Developmental psychology, 44*(3), 867.

Botero M, Macdonald SE, Miller RS (2013) Anxiety-related behavior of orphan chimpanzees (*Pan troglodytes schweinfurthii*) at Gombe National Park, Tanzania. Primates 54:21–26.

Bray, J., Emery Thompson, M., Muller, M. N., Wrangham, R. W., & Machanda, Z. P. (2018). The development of feeding behavior in wild chimpanzees (*Pan troglodytes schweinfurthii*). *American Journal of Physical Anthropology, 165*(1), 34-46.

Bretherton, I., & Bates, E. (1979). The emergence of intentional communication. New Directions for Child and *Adolescent Development, 1979*(4), 81-100.

Brigandt, I., & Love, A. C. (2010). Evolutionary novelty and the evo-devo synthesis: field notes. *Evolutionary Biology*, *37*, 93-99.

Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Machler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal, 9*(2), 378-400.

Bründl, A. C., Tkaczynski, P. J., Nohon Kohou, G., Boesch, C., Wittig, R. M., & Crockford, C. (2021). Systematic mapping of developmental milestones in wild chimpanzees. *Developmental Science, 24*(1), e12988.

Burrows, A. M., Waller, B. M., Parr, L. A., & Bonar, C. J. (2006). Muscles of facial expression in the chimpanzee (*Pan troglodytes*): descriptive, comparative and phylogenetic contexts. *Journal of anatomy*, *208*(2), 153-167.

Byrne, R. W., & Cochet, H. (2017). Where have all the (ape) gestures gone*?. Psychonomic bulletin & review, 24*, 68-71.

Byrne, R. W., Cartmill, E., Genty, E., Graham, K. E., Hobaiter, C., & Tanner, J. (2017).

    Great ape gestures: intentional communication with a rich set of innate signals.

    Animal Cognition, 20(4), 755-769.

Call J, Tomasello M (eds) (2007) The gestural communication of apes and monkeys.

    Lawrence Erlbaum Associates, New York.

Call, J. (2001). Body imitation in an enculturated orangutan (*Pongo pygmaeus*). *Cybernetics*

    & Systems, *32*(1-2), 97-119.

Call, J., & Carpenter, M. (2002). Three sources of information in social learning. *Imitation*

    in animals and artifacts, 211-228.

Cartmill, E. A., & Byrne, R. W. (2007). Orangutans modify their gestural signaling

    according to their audience's comprehension. *Current Biology, 17*(15), 1345-1348.

Cartmill, E. A., & Byrne, R. W. (2010). Semantics of primate gestures: intentional meanings

    of orangutan gestures. *Animal cognition, 13*(6), 793-804.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., ... & Stone, M.

    (1994). Animated conversation: rule-based generation of facial expression, gesture

    & spoken intonation for multiple conversational agents. In *Proceedings of the 21st*

    annual conference on Computer graphics and interactive techniques (pp. 413-420).

Chevalier-Skolnikoff, S. (1982). A cognitive analysis of facial behavior in Old World

    monkeys, apes, and human beings. In C. Snowden & C. H. Brown & M. R. Peterson

    (Eds.), Primate communication (pp. 303-368). New York: Cambridge University

    Press.

Clark, C. B. (1977). A preliminary report on weaning among chimpanzees of the Gombe

    National Park, Tanzania. *Primate bio-social development: Biological, social and*

    ecological determinants.

Clay, Z., Archbold, J., & Zuberbühler, K. (2015). Functional flexibility in wild bonobo

    vocal behaviour. *PeerJ*, *3*, e1124.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and*

    psychological measurement, 20(1), 37-46.

Crailsheim, D., Stueger, H. P., Kalcher-Sommersguter, E., & Llorente, M. (2020). Early life

    experience and alterations of group composition shape the social grooming networks

    of former pet and entertainment chimpanzees (*Pan troglodytes*). *PloS one*, *15*(1),

    e0226947.

Crockford, C., Wittig, R. M., & Zuberbühler, K. (2015). An intentional vocalization draws

    others' attention: a playback experiment with wild chimpanzees. *Animal Cognition,*

    18(3), 581-591.

Crockford, C., Wittig, R. M., & Zuberbühler, K. (2017). Vocalizing in chimpanzees is

    influenced by social-cognitive processes. *Science Advances*, *3*(11), e1701742.

Crockford, C., Wittig, R. M., Mundry, R., & Zuberbühler, K. (2012). Wild chimpanzees

    inform ignorant group members of danger. *Current Biology*, *22*(2), 142-146.

Custance, D. M., Whiten, A., & Bard, K. A. (1995). Can young chimpanzees (*Pan*

    troglodytes) imitate arbitrary actions? Hayes & Hayes (1952) revisited. *Behaviour*,

    132(11-12), 837-859.

Dafreville, M., Hobaiter, C., Guidetti, M., Sillam-Dussès, D., & Bourjade, M. (2021). Sensitivity to the communicative partner's attentional state: a developmental study on mother–infant dyads in wild chimpanzees (*Pan troglodytes schweinfurthii*). *American Journal of Primatology, 83*(12), e23339.

Darwin, C. (1872). *The descent of man, and selection in relation to sex* (Vol. 2). D. Appleton.

Davila-Ross, M., & Dezecache, G. (2021). The complexity and phylogenetic continuity of laughter and smiles in hominids. *Frontiers in Psychology*, 2065.

Davila-Ross, M., & Palagi, E. (2022). Laughter, play faces and mimicry in animals: evolution and social functions. *Philosophical Transactions of the Royal Society B, 377*(1863), 20210177.

Davila-Ross, M., Jesus, G., Osborne, J., & Bard, K. A. (2015). Chimpanzees (*Pan troglodytes*) produce the same types of 'laugh faces' when they emit laughter and when they are silent. *PloS one*, *10*(6), e0127337.

Dawkins, R., & Krebs, J. R. (1978). Animal signals: Information or manipulation? In J. R. Krebs & N. B. Davies (Eds.), The Behavioural Ecology of Animals: Evolutionary Perspectives on Animal Behavior (pp. 282-309). Blackwell.

De Ruiter, J. P. (2000). 14 The production of gesture and speech. *Language and gesture*, *2*, 284.

De Waal, F. B. (1988). The communicative repertoire of captive bonobos (*Pan paniscus*), compared to that of chimpanzees. *Behaviour*, 106(3-4), 183-251.

DeTroy, S. E., Ross, C. T., Cronin, K. A., Van Leeuwen, E. J., & Haun, D. B. (2021).

Cofeeding tolerance in chimpanzees depends on group composition: A longitudinal study across four communities. *Iscience, 24*(3), 102175.

Dezecache, G., Zuberbühler, K., Davila-Ross, M., & Dahl, C. D. (2019). Early vocal production and functional flexibility in wild infant chimpanzees. *BioRxiv*, 848770.

Dezecache, G., Zuberbühler, K., Davila-Ross, M., & Dahl, C. D. (2021). Flexibility in wild infant chimpanzee vocal behavior. *Journal of Language Evolution, 6*(1), 37-53.

Dobson, A. J., & Barnett, A. G. (2018). An introduction to generalized linear models. Chapman and Hall/CRC.

Dunbar, R. I. (1991). Functional significance of social grooming in primates. *Folia primatologica*, *57*(3), 121-131.

Dunbar, R. I. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and brain sciences*, *16*(4), 681-694.

Dunbar, R. I. (2012). Bridging the bonding gap: the transition from primates to humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1597), 1837-1846.

Dunbar, R. I. M. (1996). *Grooming, gossip, and the evolution of language*. Harvard University Press.

Dunn, J., & Shatz, M. (1989). Becoming a conversationalist despite (or because of) having an older sibling. *Child development*, 399-410.

Eaton, T., Hutton, R., Leete, J., Lieb, J., Robeson, A., & Vonk, J. (2018). Bottoms-up! Rejecting top-down human-centered approaches in comparative psychology.

International Journal of Comparative Psychology, *31*.

ELAN (Version 6.0) [Computer software]. (2020). Nijmegen: Max Planck Institute for

Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan.

Ervin-Tripp, S. (1991). Play in language development. In B. Scales, M. Almy, A.

Nicolopoulou, & S. Ervin-Tripp (Eds.), Play and the social context of development

in early care and education (pp. 84–97). Teachers College Press.

Esteve-Gibert, N., & Prieto, P. (2014). Infants temporally coordinate gesture-speech

combinations before they produce their first words. *Speech Communication, 57*, 301-

316.

Faraway, J. J. (2006). Extending the linear model with R, Texts in Statistical Science

Series.

Fedurek, P., Tkaczynski, P. J., Hobaiter, C., Zuberbühler, K., Wittig, R. M., & Crockford,

C. (2021). The function of chimpanzee greeting calls is modulated by their acoustic

variation. *Animal Behaviour*, *174*, 279-289.

Fitch, W. (2017). Empirical approaches to the study of language evolution. *Psychonomic*

bulletin & review, 24(1), 3-33.

Fleiss, J. L., Levin, B., & Paik, M. C. (1981). The measurement of interrater agreement.

Statistical methods for rates and proportions, 2(212-236), 22-23.

Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear

models: overestimated effect sizes and the winner's curse. *Behavioral Ecology and*

Sociobiology, 65(1), 47-55.

Freeberg, T. M., Dunbar, R. I., & Ord, T. J. (2012). Social complexity as a proximate and ultimate factor in communicative complexity. P*hilosophical Transactions of the Royal Society B: Biological Sciences, 367*(1597), 1785-1801.

Freeman, H. D., & Ross, S. R. (2014). The impact of atypical early histories on pet or performer chimpanzees. *PeerJ*, *2*, e579.

Fröhlich, M., & Hobaiter, C. (2018). The development of gestural communication in great-apes . *Behavioral Ecology and Sociobiology, 72*, 1-14.

Fröhlich, M., & van Schaik, C. P. (2018). The function of primate multimodal communication. *Animal Cognition, 21*(5), 619-629.

Fröhlich, M., Bartolotta, N., Fryns, C., Wagner, C., Momon, L., Jaffrezic, M., ... & van Schaik, C. P. (2021). Multicomponent and multisensory communicative acts in orang-utans may serve different functions. *Communications Biology, 4*(1), 1-13.

Fröhlich, M., Bartolotta, N., Fryns, C., Wagner, C., Momon, L., Jaffrezic, M., ... & van Schaik, C. (2020). Use and Function of Multi-Articulator Versus Multi-Sensory Acts in the Close-Range Communication of Orang-Utans.

Fröhlich, M., Bartolotta, N., Fryns, C., Wagner, C., Momon, L., Jaffrezic, M., ... & van Schaik, C. P. (2021). Multicomponent and multisensory communicative acts in orang-utans may serve different functions. *Communications Biology, 4*(1), 917.

Fröhlich, M., Müller, G., Zeiträg, C., Wittig, R. M., & Pika, S. (2017). Gestural development of chimpanzees in the wild: the impact of interactional experience. *Animal Behaviour, 134*, 271-282.

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & van Schaik, C. P. (2019).

    Multimodal communication and language origins: integrating gestures and

    vocalizations. *Biological Reviews*, *94*(5), 1809-1829.

Fröhlich, M., Wittig, R. M., & Pika, S. (2016). Should I stay or should I go? Initiation of

    joint travel in mother–infant dyads of two chimpanzee communities in the wild.

    Animal Cognition, 19, 483-500.

Fröhlich, M., Wittig, R. M., & Pika, S. (2019). The ontogeny of intentional communication

    in chimpanzees in the wild. *Developmental Science, 22*(1), e12716.

Gasser, B., & Arbib, M. (2019). A dyadic brain model of ape gestural learning, production

    and representation. *Animal cognition*, *22*, 519-534.

Genty, E. (2019). Vocal–gestural combinations in infant bonobos: new insights into signal

    functional specificity. *Animal Cognition, 22*(4), 505-518.

Genty, E., Breuer, T., Hobaiter, C., & Byrne, R. W. (2009). Gestural communication of the

    gorilla (*Gorilla gorilla*): repertoire, intentionality and possible origins. *Animal

    cognition, 12,* 527-546.

Genty, E., Clay, Z., Hobaiter, C., & Zuberbühler, K. (2014). Multi-modal use of a socially

    directed call in bonobos. *PloS one, 9*(1), e84738.

Genty, E., Neumann, C., & Zuberbühler, K. (2015). Bonobos modify communication

    signals according to recipient familiarity. *Scientific Reports, 5*(1), 1-10.

Gillespie-Lynch, K., Greenfield, P. M., Lyn, H., & Savage-Rumbaugh, S. (2014). Gestural and symbolic development among apes and humans: support for a multimodal theory of language evolution. *Frontiers in psychology, 5*, 1228.

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby) talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*(5), 339-344.

Goodall, J. (1986). The chimpanzees of Gombe: Patterns of behavior. Cambridge Mass.

Graham, K. E., Furuichi, T., & Byrne, R. W. (2020). Context, not sequence order, affects the meaning of bonobo (*Pan paniscus*) gestures. *Gesture, 19*(2-3), 335-364.

Graham, K. E., Hobaiter, C., Ounsley, J., Furuichi, T., & Byrne, R. W. (2018). Bonobo and chimpanzee gestures overlap extensively in meaning. *PLoS biology, 16*(2), e2004825.

Graham, K. E., Wilke, C., Lahiff, N. J., & Slocombe, K. E. (2020). Scratching beneath the surface: intentionality in great ape signal production. *Philosophical Transactions of the Royal Society B*, *375*(1789), 20180403.

Grice, H. P. (1969). Utterer's meaning and intention. *The philosophical review*, *78*(2), 147-177.

Grice, H. P. (1969). Utterer's meaning and intention. *The philosophical review*, *78*(2), 147-177.

Gros-Louis, J., & Wu, Z. (2012). Twelve-month-olds' vocal production during pointing in naturalistic interactions: Sensitivity to parents' attention and responses. *Infant Behavior and Development, 35*(4), 773.

Halina, M., Rossano, F., & Tomasello, M. (2013). The ontogenetic ritualization of bonobo gestures. Animal cognition, 16, 653-666.

Hamada, Y., & Udono, T. (2002). Longitudinal analysis of length growth in the chimpanzee (*Pan troglodytes*). *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists, 118*(3), 268-284.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve?. *science*, *298*(5598), 1569-1579.

Hawkes, K., O'Connell, J. F., Jones, N. B., Alvarez, H., & Charnov, E. L. (1998). Grandmothering, menopause, and the evolution of human life histories. *Proceedings of the National Academy of Sciences, 95*(3), 1336-1339.

Heesen, R., Sievers, C., Gruber, T., & Clay, Z. (2022). Primate communication: Affective, intentional, or both?. *Primate cognitive studies*.

Henzi, S. P., Lycett, J. E., & Piper, S. E. (1997). Fission and troop size in a mountain baboon population. *Animal behaviour*, *53*(3), 525-535.

Higham, J. P., & Hebets, E. A. (2013). An introduction to multimodal communication. *Behavioral Ecology and Sociobiology, 67*, 1381-1388.

Hobaiter, C., & Byrne, R. W. (2011). Serial gesturing by wild chimpanzees: its nature and function for communication. *Animal cognition, 14*, 827-838.

Hobaiter, C., & Byrne, R. W. (2011). The gestural repertoire of the wild chimpanzee. *Animal cognition, 14*, 745-767.

Hobaiter, C., & Byrne, R. W. (2014). The meanings of chimpanzee gestures. *Current Biology*, *24*(14), 1596-1600.

Hobaiter, C., Byrne, R. W., & Zuberbühler, K. (2017). Wild chimpanzees' use of single and combined vocal and gestural signals. *Behavioral Ecology and Sociobiology, 71*(6), 1-13.

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*(3), 88-97.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental review, 26*(1), 55-88.

Hoff, E. (2013). Language development. Cengage Learning..

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences, 23*(8), 639-652.

Holler, J., Kendrick, K. H., & Levinson, S. C. (2018). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic bulletin & review,* 25, 1900-1908.

Holler, J., Kendrick, K. H., & Levinson, S. C. (2018). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic bulletin & review*, *25*, 1900-1908.

Hooper, P. L., Gurven, M., Winking, J., & Kaplan, H. S. (2015). Inclusive fitness and

    differential productivity across the life course determine intergenerational transfers

    in a small-scale human society. *Proceedings of the Royal Society B: Biological*

    Sciences, 282(1803), 20142808.

Hrdy, S. B. (2009). *Mothers and others: The evolutionary origins of mutual understanding*.

    Harvard University Press.

Humle, T., Maisels, F., Oates, J. F., Plumptre, A., & Williamson, E. A. (2016). *Pan*

    troglodytes (errata version published in 2018).

Humle, T., Maisels, F., Oates, J. F., Plumptre, A., & Williamson, E. A. (2016). *Pan*

    troglodytes (errata version published in 2018). *The IUCN Red List of Threatened*

    Species. Accessed on 19 May 2023.

Humphrey, L. T. (2010, June). Weaning behaviour in human evolution. In *Seminars in cell*

    & developmental biology (Vol. 21, No. 4, pp. 453-461). Academic Press.

Igualada, A., Bosch, L., & Prieto, P. (2015). Language development at 18 months is related

    to multimodal communicative strategies at 12 months. *Infant Behavior and*

    Development, 39, 42-52.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between

    motor development and language development. *Journal of child language, 37*(2),

    229-261.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language

    development. *Psychological Science, 16*(5), 367-371.

Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness studies, 6*(11-12), 19-40.

Iverson, J. M., Capirci, O., Longobardi, E., & Caselli, M. C. (1999). Gesturing in mother-child interactions. *Cognitive development, 14*(1), 57-75.

Izumi, A., & Kojima, S. (2004). Matching vocalizations to vocalizing faces in a chimpanzee (*Pan troglodytes*). *Animal cognition*, *7*, 179-184.

Kalcher-Sommersguter, E., Preuschoft, S., Franz-Schaider, C., Hemelrijk, C. K., Crailsheim, K., & Massen, J. J. (2015). Early maternal loss affects social integration of chimpanzees throughout their lifetime. *Scientific Reports*, *5*(1), 16439.

Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2011). Transition from crawling to walking and infants' actions with objects and people. *Child development, 82*(4), 1199-1209.

Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2014). Crawling and walking infants elicit different verbal responses from mothers. *Developmental science, 17*(3), 388-395.

Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional expression: Advances in basic emotion theory. *Journal of nonverbal behavior, 43*(2), 133-160.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. *The relationship of verbal and nonverbal communication*, *25*(1980), 207-227.

Kendon, A. (2004). Gesture: Visible action as utterance. Cambridge University Press.

Kita, S. (2000). How representational gestures help speaking. *Language and gesture*, *1*, 162-185.

Kojima, S. (2001). Early vocal development in a chimpanzee infant. Primate origins of
human cognition and behavior, 190-196.

Kojima, S. (2008). Early Vocal Development in a Chimpanzee Infant. In: Matsuzawa, T.
(eds) Primate Origins of Human Cognition and Behavior. Springer, Tokyo.

Kramer, K. L., & Ellison, P. T. (2010). Pooled energy budgets: Resituating human energy-
allocation trade-offs. *Evolutionary Anthropology: Issues, News, and Reviews, 19*(4),
136-147.

Lameira, A. R., & Shumaker, R. W. (2019). Orangutans show active voicing through a
membranophone. *Scientific Reports*, *9*(1), 1-6.

Laporte, M. N., & Zuberbühler, K. (2011). The development of a greeting signal in wild
chimpanzees. *Developmental Science, 14*(5), 1220-1234.

Leavens, D. A., & Bard, K. A. (2011). Environmental influences on joint attention in great-
apes : implications for human cognition. *Journal of Cognitive Education and
Psychology*, *10*(1), 9.

Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2005). Intentionality as measured in the
persistence and elaboration of communication by chimpanzees (*Pan troglodytes*).
Child development, *76*(1), 291-306.

Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2010). Multimodal communication by
captive chimpanzees (*Pan troglodytes*). *Animal cognition, 13,* 33-40.

Leigh, S. R. (2004). Brain growth, life history, and cognition in primate and human evolution. *American Journal of Primatology: Official Journal of the American Society of Primatologists, 62*(3), 139-164.

Leigh, S. R., & Shea, B. T. (1996). Ontogeny of body size variation in African apes. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists, 99*(1), 43-65.

Lembeck M (2015) Parenting behaviour in the first year of life: a cross-species comparison of humans, chimpanzees, bonobos and gibbons. Ph.D. Dissertation, Freie Universität Berlin, Berlin

Leroux, M., Bosshard, A. B., Chandia, B., Manser, A., Zuberbühler, K., & Townsend, S. W. (2021). Chimpanzees combine pant hoots with food calls into larger structures. *Animal Behaviour*, *179*, 41-50.

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends in cognitive sciences*, *20*(1), 6-14.

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651), 20130302.

Liebal, K., & Oña, L. (2018). Mind the gap–moving beyond the dichotomy between intentional gestures and emotional facial and vocal signals of nonhuman primates. *Interaction Studies, 19*(1-2), 121-135.

Liebal, K., Call, J., & Tomasello, M. (2004). Use of gesture sequences in chimpanzees. *American Journal of Primatology: Official Journal of the American Society of Primatologists, 64*(4), 377-396.

Liebal, K., Schneider, C., & Errson-Lembeck, M. (2019). How primates acquire their gestures: evaluating current theories and evidence. *Animal Cognition*, *22*, 473-486.

Liebal, K., Slocombe, K. E., & Waller, B. M. (2022). The language void 10 years on: multimodal primate communication research is still uncommon. *Ethology Ecology & Evolution*, *34*(3), 274-287.

Liebal, K., Waller, B. M., Slocombe, K. E., & Burrows, A. M. (2014). Primate communication: a multimodal approach. Cambridge University Press.

Lieberman, P., & McCarthy, R. C. (1975). The evolution of speech and language. *The role of speech in language*, 83-106.

Liszkowski, U. (2011). Three lines in the emergence of prelinguistic communication and social cognition. *Journal of cognitive education and psychology, 10*(1), 32-43.

Liszkowski, U., Albrecht, K., Carpenter, M., & Tomasello, M. (2008). Infants' visual and auditory communication when a partner is or is not visually attending. *Infant Behavior and Development, 31*(2), 157-167.

Llorente, M., Riba, D., Ballesta, S., Feliu, O., & Rostán, C. (2015). Rehabilitation and socialization of chimpanzees (*Pan troglodytes*) used for entertainment and as pets: an 8-year study at Fundació Mona. *International Journal of Primatology*, *36*, 605-624.

Locke, J. L., & Bogin, B. (2006). Language and life history: A new perspective on the development and evolution of human language. *Behavioral and Brain Sciences*, *29*(3), 259-280.

Lonsdorf, E. V. (2006). What is the role of mothers in the acquisition of termite-fishing behaviors in wild chimpanzees (*Pan troglodytes schweinfurthii*)?. *Animal cognition*, *9*, 36-46.

Lonsdorf, E. V. (2012). The role of mothers in the development of complex skills in chimpanzees. *Building babies: Primate development in proximate and ultimate perspective*, 303-318.

Lopez, M., Ruiz, M. O., Rovnaghi, C. R., Tam, G. K., Hiscox, J., Gotlib, I. H., ... & Anand, K. J. (2021). The social ecology of childhood and early life adversity. *Pediatric research*, *89*(2), 353-367.

Luef, E. M., & Pika, S. (2017). Reciprocal greeting in chimpanzees (*Pan troglodytes*) at the Ngogo community. *Journal of Neurolinguistics, 43*, 263-273.

Maestripieri, D. (2009). 12. Maternal Influences on Offspring Growth, Reproduction, and Behavior in Primates. In Maternal effects in mammals (pp. 256-291). University of Chicago Press.

Maestripieri, D., & Call, J. (1996). Mother-infant communication in primates. *Advances in the Study of Behavior, 25*, 613-642.

Maggi, S., Irwin, L. J., Siddiqi, A., & Hertzman, C. (2010). The social determinants of early
child development: an overview. *Journal of paediatrics and child health*, *46*(11),
627-635.

Marstaller, L., & Burianová, H. (2015). A common functional neural network for overt
production of speech and gesture. *Neuroscience*, 284, 29-41.

Marzke, M. W., Young, D. L., Hawkey, D. E., Su, S. M., Fritz, J., & Alford, P. L. (1996).
Comparative analysis of weight gain, hand/wrist maturation; and dental emergence
rates in chimpanzees aged 0–24 months from varying captive environments.
*American Journal of Physical Anthropology: The Official Publication of the
American Association of Physical Anthropologists, 99*(1), 175-190.

Massaro, D. W. & Egan, P. B. (1996). Perceiving affect from the voice and the face.
Psychonomic Bulletin & Review 3, 215– 221.

Massaro, D. W. (1998). Perceiving Talking Faces: From Speech Perception to A Behavioral
Principle. MIT Press, Cambridge

McCartney, K. (1984). Effect of quality of day care environment on children's language
development. *Developmental psychology, 20*(2), 244.

McCullagh, P. and Nelder, J.A. (1989) Generalized Linear Models. 2nd Edition, Chapman
and Hall, London.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–
748.

McNeill, D. (1992). Hand and Mind: What Gestures Reveal about Thought. University of Chicago Press.

Messinger, D. S., & Fogel, A. (1998). Give and Take: The Development of Conventional Infant Gestures. *Merrill-Palmer Quarterly*, *44*(4), 566–590. 566-590.

Mitani, J. C. (2009). Male chimpanzees form enduring and equitable social bonds. *Animal Behaviour*, *77*(3), 633-640.

Mitani, J. C., Merriwether, D. A., & Zhang, C. (2000). Male affiliation, cooperation and kinship in wild chimpanzees. *Animal behaviour*, *59*(4), 885-893.

Mizuno, Y., Takeshita, H., & Matsuzawa, T. (2006). Behavior of infant chimpanzees during the night in the first 4 months of life: Smiling and suckling in relation to behavioral state. *Infancy*, *9*(2), 221-240.

Morgenstern, A. (2014). Children's multimodal language development. *Manual of language acquisition*, 123-142.

Morgenstern, A. (2023). Children's multimodal language development from an interactional, usage-based, and cognitive perspective. *Wiley Interdisciplinary Reviews: Cognitive Science, 14*(2), e1631.

Müller, G. B. (2007). Evo–devo: extending the evolutionary synthesis. *Nature reviews genetics, 8*(12), 943-949.

Mundry, R. (2014). Statistical issues and assumptions of phylogenetic generalized least squares. Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice, 131-153.

Murillo, E., & Belinchón, M. (2012). Gestural-vocal coordination: Longitudinal changes and predictive value on early lexical development. *Gesture, 12*(1), 16-39.

Murillo, E., & Casla, M. (2021). Multimodal representational gestures in the transition to multi-word productions. *Infancy, 26*(1), 104-122.

Murillo, E., Ortega, C., Otones, A., Rujas, I., & Casla, M. (2018). Changes in the synchrony of multimodal communication in early language development. *Journal of speech, language, and hearing research*, *61*(9), 2235-2245.

Murphy, C. M., & Messer, D. J. (1977). Mothers, infants and pointing: A study of gesture. In H. R. Schaffer (Ed.), Studies in mother-infant interaction (pp. 325-354). London: Academic.

Myowa-Yamakoshi, M., & Matsuzawa, T. (1999). Factors influencing imitation of manipulatory actions in chimpanzees (*Pan troglodytes*). *Journal of comparative psychology*, *113*(2), 128.

Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M., & Matsuzawa, T. (2004). Imitation in neonatal chimpanzees (*Pan troglodytes*). *Developmental science*, *7*(4), 437-442.

National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network. (2000). Characteristics and quality of child care for toddlers and preschoolers. *Applied Developmental Science, 4*(3), 116-135.

Nishida, T. (1968). The social group of wild chimpanzees in the Mahali Mountains. *Primates*, *9*, 167-224.

Nishida, T. (2011). *Chimpanzees of the lakeshore: natural history and culture at Mahale*. Cambridge University Press.

Nishida, T., Kano, T., Goodall, J., McGrew, W. C., & Nakamura, M. (1999). Ethogram and ethnography of Mahale chimpanzees. *Anthropological Science, 107*(2), 141-188.

Nishida, T., Takasaki, H., & Takahata, Y. (1990). Demography and reproductive profiles. *The chimpanzees of the Mahale mountains: Sexual and life history strategies*, 63-97.

Noroozi, F., Corneanu, C. A., Kamińska, D., Sapiński, T., Escalera, S., & Anbarjafari, G. (2018). Survey on emotional body gesture recognition. *IEEE transactions on affective computing*, *12*(2), 505-523.

Noroozi, F., Corneanu, C. A., Kamińska, D., Sapiński, T., Escalera, S., & Anbarjafari, G. (2018). Survey on emotional body gesture recognition. *IEEE transactions on affective computing*, *12*(2), 505-523.

Oller, D. K. (2012). Underpinnings for a theory of communicative evolution. In D. K. Oller, & U. Griebel (Eds.), Evolution of communication systems: A comparative approach (pp. 49–67). MIT Press.

Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences*, *110*(16), 6318-6323.

Oller, D. K., Griebel, U., Iyer, S. N., Jhang, Y., Warlaumont, A. S., Dale, R., & Call, J. (2019). Language origins viewed in spontaneous and interactive vocal rates of human and bonobo infants. *Frontiers in psychology*, *10*, 729.

Oller, D.K. (2000). The Emergence of the Speech Capacity (1st ed.). Psychology Press.

Oña, L. S., Sandler, W., & Liebal, K. (2019). A stepping stone to compositionality in chimpanzee communication. *PeerJ, 7*, e7623.

Oshima-Takane, Y., Goodz, E., & Derevensky, J. L. (1996). Birth order effects on early language development: Do secondborn children learn from overheard speech? *Child development, 67*(2), 621-634.

Özçalışkan, Ş., & Goldin-Meadow, S. (2011). Is there an iconic gesture spurt at 26 months. *Integrating gestures: The interdisciplinary nature of gesture. Amsterdam, NL: John Benjamins.*

Palagi, E., Cordoni, G., & Borgognini Tarli, S. (2004). Immediate and delayed benefits of play behaviour: new evidence from chimpanzees (*Pan troglodytes*). *Ethology, 110*(12), 949-962.

Palagi, E., Norscia, I., Pressi, S., & Cordoni, G. (2018). Facial mimicry and play: A comparative study in chimpanzees and gorillas. *Emotion*, *19*(4), 665.

Parker, J. G., & Asher, S. R. (1987). Peer relations and later personal adjustment: Are low-accepted children at risk? *Psychological bulletin, 102*(3), 357.

Parr, L. A. (2004). Perceptual biases for multimodal cues in chimpanzee (*Pan troglodytes*) affect recognition. *Animal cognition*, *7*, 171-178.

Parr, L. A., Cohen, M., & Waal, F. d. (2005). Influence of social context on the use of blended and graded facial displays in chimpanzees. *International Journal of Primatology, 26*(1), 73-103.

Parr, L. A., Waller, B. M., Vick, S. J., & Bard, K. A. (2007). Classifying chimpanzee facial expressions using muscle action. *Emotion, 7*(1), 172.

Partan, S. R. (2013). Ten unanswered questions in multimodal communication. *Behavioral Ecology and Sociobiology, 67*(9), 1523-1539.

Partan, S. R., & Marler, P. (2005). Issues in the classification of multimodal communication signals. *The American Naturalist, 166*(2), 231-245.

Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science, 283*(5406), 1272-1273.

PASA. (2023). PASA. Retrieved 12 May 2023, from https://pasa.org

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & Yazejian, N. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child development, 72*(5), 1534-1553.

Pellegrini, A. D., Galda, L., Flor, D., Bartini, M., & Charak, D. (1997). Close relationships, individual differences, and early literacy learning. *Journal of experimental child psychology, 67*(3), 409-422.

Pepper, J. W., Mitani, J. C., & Watts, D. P. (1999). General gregariousness and specific social preferences among wild chimpanzees. *International Journal of Primatology*, *20*(5), 613.

Pika, S., & Fröhlich, M. (2019). Gestural acquisition in great apes : the social negotiation hypothesis. *Animal Cognition*, *22*(4), 551-565.

Pika, S., Liebal, K., & Tomasello, M. (2003). Gestural communication in young gorillas (*Gorilla gorilla*): gestural repertoire, learning, and use. *American Journal of Primatology: Official Journal of the American Society of Primatologists, 60*(3), 95-111.

Pilbeam, D. (1996). Genetic and morphological records of the Hominoidea and hominid origins: a synthesis. *Molecular phylogenetics and evolution*, *5*(1), 155-168.

Pilbeam, D., & Young, N. (2004). Hominoid evolution: synthesizing disparate data. *Comptes Rendus Palevol*, *3*(4), 305-321.

Pillas, D., Marmot, M., Naicker, K., Goldblatt, P., Morrison, J., & Pikhart, H. (2014). Social inequalities in early childhood health and development: a European-wide systematic review. *Pediatric research*, *76*(5), 418-424.

Plooij, F. X. (1978). Some basic traits of language in wild chimpanzees? In (A. Lock, Ed.) Action, Gesture and Symbol.

Plooij, F. X. (1979). How wild chimpanzee babies trigger the onset of mother-infant play. *Before speech,* 223-43.

Plooij, F. X. (1984). The behavioral development of free-living chimpanzee babies and infants. Monographs on infancy.

Pollick, A. S., & De Waal, F. B. (2007). Ape gestures and language evolution. *Proceedings of the National Academy of Sciences, 104*(19), 8184-8189.

Pusey, A. E. (1983). Mother-offspring relationships in chimpanzees after weaning. *Animal Behaviour, 31*(2), 363-377.

Pusey, A. E. (1990). Behavioural changes at adolescence in chimpanzees. *Behaviour, 115*(3-4), 203-246.

Quinn, G. P., & Keough, M. J. (2002). Experimental design and data analysis for biologists. Cambridge university press.

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *http://www.R-project. org/*.

Reiches, M. W., Ellison, P. T., Lipson, S. F., Sharrock, K. C., Gardiner, E., & Duncan, L. G. (2009). Pooled energy budget and human life history. *American Journal of Human Biology: The Official Journal of the Human Biology Association, 21*(4), 421-429.

Renzi, D. T., Romberg, A. R., Bolger, D. J., & Newman, R. S. (2017). Two minds are better than one: Cooperative communication as a new framework for understanding infant language learning. *Translational Issues in Psychological Science, 3*(1), 19.

Roberts, A. I., & Roberts, S. G. B. (2016). Wild chimpanzees modify modality of gestures according to the strength of social bonds and personal network size. *Scientific reports*, *6*(1), 1-13.

Roberts, A. I., Vick, S. J., & Buchanan-Smith, H. M. (2012). Usage and comprehension of manual gestures in wild chimpanzees. *Animal Behaviour*, *84*(2), 459-470.

Roberts, S. G. B., & Roberts, A. I. (2019). Visual attention, indicative gestures, and calls accompanying gestural communication are associated with sociality in wild chimpanzees (*Pan troglodytes schweinfurthii*). *Journal of Comparative Psychology*, *133*(1), 56.

Roberts, S. G., & Roberts, A. I. (2016). Social brain hypothesis: Vocal and gesture networks of wild chimpanzees. *Frontiers in psychology*, 1756.

Ron, T., & McGrew, W. (1988). Ecological assessment for a chimpanzee rehabilitation project in Northern Zambia. *Primate Conservation, 9*, 37-41.

Rosenberg, K. R. (2021). The evolution of human infancy: Why it helps to be helpless. *Annual Review of Anthropology, 50*, 423-440.

Ross, K. M., Bard, K. A., & Matsuzawa, T. (2014). Playful expressions of one-year-old chimpanzee infants in social and solitary play contexts. *Frontiers in Psychology*, *5*, 741.

Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. *Animal behaviour*, *58*(5), 921-931.

Russell, J. L., Lyn, H., Schaeffer, J. A., & Hopkins, W. D. (2011). The role of socio-communicative rearing environments in the development of social and physical cognition in apes. *Developmental science*, *14*(6), 1459-1470.

Rutter, M., Beckett, C., Castle, J., Colvert, E., Kreppner, J., Mehta, M., ... & Sonuga-Barke, E. (2007). Effects of profound early institutional deprivation: An overview of findings from a UK longitudinal study of Romanian adoptees. *European Journal of Developmental Psychology*, *4*(3), 332-350.

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., ... & Cohen, D. (2013). Motherese in interaction: at the cross-road of emotion and cognition? (A systematic review). *PloS one, 8*(10), e78103.

Schaffer, H. R., Hepburn, A., & Collis, G. M. (1983). Verbal and nonverbal aspects of mothers' directives. Journal of *Child Language, 10*(2), 337-355.

Scheider, L., Waller, B. M., Oña, L., Burrows, A. M., & Liebal, K. (2016). Social use of facial expressions in hylobatids. *PloS one*, *11*(3), e0151733.

Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., & Slocombe, K. E. (2013). Chimpanzee alarm call production meets key criteria for intentionality. *PloS one*, *8*(10), e76674.

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution, 1*(2), 103-113.

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: overconfident estimates in mixed models. *Behavioral ecology, 20*(2), 416-420.

Schmidt, L. (1996) Psykosociale konsekvenser af infertilitet og behandling [Psychosocial consequences of infertility and treatment]. Copenhagen: FADL Press.

Schneider, C., Call, J., & Liebal, K. (2012). Onset and early use of gestural communication in nonhuman great apes . *American Journal of Primatology, 74*(2), 102-113.

Schneider, C., Call, J., & Liebal, K. (2012). What role do mothers play in the gestural acquisition of bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*)? *International Journal of Primatology, 33*(1), 246-262.

Schneider, C., Liebal, K., & Call, J. (2017). "Giving" and "responding" differences in gestural communication between nonhuman great ape mothers and infants. *Developmental psychobiology, 59*(3), 303-313.

Slocombe, K. E., & Zuberbühler, K. (2005). Functionally referential communication in a chimpanzee. *Current Biology*, *15*(19), 1779-1784.

Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences*, *104*(43), 17228-17233.

Slocombe, K. E., & Zuberbühler, K. (2010). Vocal communication in chimpanzees. *The mind of the chimpanzee: ecological and experimental perspectives*, 192-207.

Slocombe, K. E., Waller, B. M., & Liebal, K. (2011). The language void: the need for multimodality in primate communication research. *Animal Behaviour, 81*(5), 919-924.

Smith, C. L., & Evans, C. S. (2013). A new heuristic for capturing the complexity of multimodal signals. Behavioral Ecology and Sociobiology, 67, 1389-1398.

Smith, W. J. (1965). Message, meaning, and context in ethology. *The American Naturalist, 99*(908), 405-409.

Smith, W. J. (1977). The behavior of communicating: an ethological approach. Harvard university press.

Snowdon, C. T., & Hausberger, M. (Eds.). (1997). *Social influences on vocal development*. Cambridge University Press.

Soldati, A., Muhumuza, G., Dezecache, G., Fedurek, P., Taylor, D., Call, J., & Zuberbühler, K. (2022). The ontogeny of vocal sequences: Insights from a newborn wild chimpanzee (*Pan troglodytes schweinfurthii*). *International Journal of Primatology*, 1-24.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587-10592.

Subiaul, F. (2016). What's special about human imitation? A comparison with enculturated apes. *Behavioral Sciences*, *6*(3), 13.

Taglialatela, J. P., Russell, J. L., Pope, S. M., Morton, T., Bogart, S., Reamer, L. A., Schapiro, S. J., & Hopkins, W. D. (2015). Multimodal communication in chimpanzees. *American journal of primatology, 77*(11), 1143-1148.

Tamis-LeMonda, C. S., Cristofaro, T. N., Rodriguez, E. T., & Bornstein, M. H. (2006). Early Language Development: Social Influences in the First Years of Life.

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science, 23*(2), 121-126.

Tamis-Lemonda, C. S., Tafuro, L., Kuchirko, Y., Song, L., & Kahana-Kalman, R. (2013). Mothers' responsiveness and child development: A focus on mother-infant interactions in low-income, ethnically diverse families. In Biennial Meeting of the Society for Research on Child Development, Seattle, WA.

Tarr, B., Launay, J., & Dunbar, R. I. (2016). Silent disco: dancing in synchrony leads to elevated pain thresholds and social closeness. *Evolution and Human Behavior*, *37*(5), 343-349.

Taylor, D., Clay, Z., Dahl, C. D., Zuberbühler, K., Davila-Ross, M., & Dezecache, G. (2022). Vocal functional flexibility: what it is and why it matters. *Animal Behaviour, 186*, 93-100.

Taylor, D., Dezecache, G., & Davila-Ross, M. (2021). Filling in the gaps: acoustic gradation increases in the vocal ontogeny of chimpanzees (*Pan troglodytes*). *American Journal of Primatology, 83*(5), e23249.

Taylor, D., Gustafsson, E., Dezecache, G., & Davila-Ross, M. (2022). Directedness and engagement in chimpanzee vocal ontogeny. *Developmental Science*, e13334.

Ter Bekke, M., Drijvers, L., & Holler, J. (2020). The predictive potential of hand gestures during conversation: An investigation of the timing of gestures in relation to speech. *PsyArXiv Preprints.*

Thompson, M. E., Muller, M. N., & Wrangham, R. W. (2012). The energetics of lactation and the return to fecundity in wild chimpanzees. *Behavioral Ecology, 23*(6), 1234-1241.

Tomasello, M., & Call, J. (2007). Ape gestures and the origins of language. In J. Call & M. Tomasello (Eds.), The gestural communication of apes and monkeys (pp. 221–239). Taylor & Francis Group/Lawrence Erlbaum Associates.

Tomasello, M., Call, J., Nagell, K., Olguin, R., & Carpenter, M. (1994). The learning and use of gestural signals by young chimpanzees: A trans-generational study. *Primates, 35,* 137-154.

Tomasello, M., Call, J., Warren, J., Frost, G. T., Carpenter, M., & Nagell, K. (1997). The
ontogeny of chimpanzee gestural signals: a comparison across groups andtomasell
generations. *Evolution of communication, 1*(2), 223-259.

Tomasello, M., George, B. L., Kruger, A. C., Jeffrey, M., & Evans, A. (1985). The
development of gestural communication in young chimpanzees. *Journal of Human
Evolution, 14*(2), 175-186.

Tomasello, M., Gust, D., & Frost, G. T. (1989). A longitudinal investigation of gestural
communication in young chimpanzees. Primates, 30, 35-50.

Tomasello, M., Savage-Rumbaugh, S., & Kruger, A. C. (1993). Imitative learning of actions
on objects by children, chimpanzees, and enculturated chimpanzees. *Child
development*, *64*(6), 1688-1705.

Townsend, S. W., Koski, S. E., Byrne, R. W., Slocombe, K. E., Bickel, B., Boeckle, M.,
Braga Goncalves, I., Burkart, J. M., Flower, T., & Gaunet, F. (2017). Exorcising G
rice's ghost: An empirical approach to studying intentional communication in
animals. *Biological Reviews, 92*(3), 1427-1433.

Townsend, S. W., Watson, S. K., Slocombe, K. E., Hopper, L. M., & Ross, S. R. (2020).
Flexibility in Great Ape Vocal Production.

Ueno, A., & Matsuzawa, T. (2004). Food transfer between chimpanzee mothers and their
infants. *Primates*, *45*, 231-239.

Valbonesi, L., Ansari, R., McNeill, D., Quek, F., Duncan, S., McCullough, K. E., & Bryll,
R. (2002, September). Multimodal signal analysis of prosody and hand motion:

Temporal correlation of speech and gestures. In *2002 11th European Signal Processing Conference* (pp. 1-4). IEEE.

Van de Pol, M., & Wright, J. (2009). A simple method for distinguishing within-versus between-subject effects using mixed models. *Animal behaviour, 77*(3), 753-758.

van de Rijt-Plooij, H. H., & Plooij, F. X. (1987). Growing independence, conflict and learning in mother-infant relations in free-ranging chimpanzees. *Behaviour, 101*(1-3), 1-86.

van Lawick-Goodall, J. (1967). Mother-offspring relationships in free-ranging chimpanzees. In D. Morris (Ed.), Primate ethology (pp. 287–346). London: Weidenfeld and Nicolson

van Lawick-Goodall, J. (1968). A preliminary report on expressive movements and communication in the Gombe stream chimpanzees. In P. C. Jay (Ed.), Primates, studies in adaptation and variability (pp.313–374). New York: Holt, Rinehart & Winston.pa

van Leeuwen, E. J., Bruinstroop, B. M., & Haun, D. B. (2022). Early Trauma Leaves No Social Signature in Sanctuary-Housed Chimpanzees (*Pan troglodytes*). *Animals*, *13*(1), 49.

Van Leeuwen, E. J., Cronin, K. A., & Haun, D. B. (2018). Population-specific social dynamics in chimpanzees. *Proceedings of the National Academy of Sciences, 115*(45), 11393-11400.

Van Leeuwen, E. J., Mulenga, I. C., & Chidester, D. L. (2014). Early social deprivation negatively affects social skill acquisition in chimpanzees (*Pan troglodytes*). *Animal Cognition*, *17*, 407-414.

Van Schaik, C. P. (2016). The primate origins of human nature. John Wiley & Sons.

Vatikiotis-Bateson, E., Munhall, K. G., Kasahara, Y., Garcia, F., & Yehia, H. (1996, October). Characterizing audiovisual information during speech. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96* (Vol. 3, pp. 1485-1488). IEEE.

Vick, S.-J., Waller, B. M., Parr, L. A., Smith Pasqualini, M. C., & Bard, K. A. (2007). A cross-species comparison of facial morphology and movement in humans and chimpanzees using the facial action coding system (FACS). *Journal of nonverbal behavior, 31*(1), 1-20.

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651), 20130292.

Vihman, M. M. (2014). Phonological development: The first two years. John Wiley & Sons.

Walker, K. K., Walker, C. S., Goodall, J., & Pusey, A. E. (2018). Maturation is prolonged and variable in female chimpanzees. Journal of human evolution, 114, 131-140.

Waller, B. M., & Dunbar, R. I. (2005). Differential behavioural effects of silent bared teeth display and relaxed open mouth display in chimpanzees (*Pan troglodytes*). *Ethology, 111*(2), 129-142.

Waller, B. M., Caeiro, C. C., & Davila-Ross, M. (2015). Orangutans modify facial displays depending on recipient attention. *PeerJ*, *3*, e827.

Waller, B. M., Liebal, K., Burrows, A. M., & Slocombe, K. E. (2013). How can a multimodal approach to primate communication help us understand the evolution of communication? *Evolutionary Psychology, 11*(3), 147470491301100305.

Waller, B. M., Whitehouse, J., & Micheletta, J. (2017). Rethinking primate facial expression: A predictive framework. *Neuroscience & Biobehavioral Reviews*, *82*, 13-21.

Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., ... & Slocombe, K. E. (2015). Vocal learning in the functionally referential food grunts of chimpanzees. *Current Biology*, *25*(4), 495-499.

Watts, D. P. (2006). Conflict resolution in chimpanzees and the valuable-relationships hypothesis. *International Journal of Primatology*, *27*, 1337-1364.

Wetherby, A. M., & Prizant, B. M. (1989). The expression of communicative intent: Assessment guidelines. *Seminars in Speech and Language*, *10*(1), 77-91.

Wilke, C., Kavanagh, E., Donnellan, E., Waller, B. M., Machanda, Z. P., & Slocombe, K. E. (2017). Production of and responses to unimodal and multimodal signals in wild chimpanzees, *Pan troglodytes schweinfurthii. Animal Behaviour, 123*, 305-316.

Williams, J. M., Pusey, A. E., Carlis, J. V., Farm, B. P., & Goodall, J. (2002). Female competition and male territorial behaviour influence female chimpanzees' ranging patterns. *Animal Behaviour*, *63*(2), 347-360.

Wrangham, R., & Pilbeam, D. (2001). African apes as time machines. *All Apes Great and Small: Volume 1: African Apes*, 5-17.

Wu, Z., & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language, 34*(1), 72-90.

Yale, M. E., Messinger, D. S., Cobo-Lewis, A. B., & Delgado, C. F. (2003). The temporal coordination of early infant communication. *Developmental psychology*, *39*(5), 815.

Yale, M. E., Messinger, D. S., Cobo-Lewis, A. B., Oller, D. K., & Eilers, R. E. (1999). An event-based analysis of the coordination of early infant vocalizations and facial actions. *Developmental Psychology*, *35*(2), 505.

Zuberbühler, K. (2020). Syntax and compositionality in animal communication. *Philosophical Transactions of the Royal Society B*, *375*(1789), 20190062.

Appendices

Appendix 1

Appendix 1.1

Table A1.1. Full group size information in 2017 and 2021 field periods

Observation year	Group number		
	1	2	4
2017	25	52	12
2021	28	58	NA

NA: individuals from group 4 were not observed in 2021.

Table A1.2. The number of each signal type coded within each behavioural context

Context	Signal type			Total (no. of individuals, individual mean ± SD)
	Unimodal	Multimodal combination	Other	
Feeding	184	50	68	302 (27, 10.79 ± 9.92)
Play	701	174	85	960 (27, 34.29 ± 33.61)
Aggression	54	74	73	201 (22, 7.18 ± 10.98)
Access*	12	8	3	23 (8, 0.82 ± 1.47)
Affiliation*	52	15	12	79 (22, 2.82 ± 2.69)
Grooming*	31	2	10	43 (13, 1.54 ± 3.32)
Rest*	20	4	8	32 (16, 1.14 ± 1.38)
Sexual*	23	3	3	29 (6, 1.04 ± 3.47)
Travel*	7	2	4	13 (8, 0.46 ± 0.82)

The 'other' signal type includes fixed multimodal, unimodal, and multimodal sequences.

*Included in 'other' context category.

Appendix 1.2

The focus of this study was to study patterns of unimodal and multimodal signal production

in immature chimpanzees. While fixed multimodal signals occur simultaneously by default, for consistency with recent research on multimodal signal production in great-apes (e.g., Luef & Pika, 2017; Wilke et al., 2017; Fröhlich et al. 2019; Oña et al., 2019) we considered free multimodal signal combinations (henceforth multimodal combinations) to be temporally overlapping combinations of facial expressions, gestures, and vocalizations. However, in the past, others have allowed a time gap of up to 10 s between the production of the different modalities comprising a multimodal combination (Pollick & de Waal, 2007). Genty (2019) included simultaneously produced signals (call + gesture) as a multimodal combination but also included instances when one modality was produced <1 s after the previous one. However, in the same paper a unimodal (call) sequence was defined as individual call units produced <1 s apart, in line with other unimodal studies where a sequence was defined under the same criterion (e.g., Hobaiter & Byrne, 2011b; Genty et al., 2014; Graham et al., 2018).

For consistency with these unimodal studies involving signal sequences, we wanted to separate overlapping multimodal combinations from the production of multiple signals that temporally speaking would fall under the definition of a 'sequence' elsewhere. When we encountered signals of multiple modalities (e.g., call + gesture) produced within 1 s of each other (see also Genty et al., 2014), we differentiated these as a multimodal sequence. Unimodal sequences were then defined in the same way as multimodal sequences, but signals were from one modality (e.g., gesture), as seen in previous unimodal research. Sequences involving fixed multimodal signals, for example scream face + scream vocalization, were included as multimodal sequences. We did not consider multimodal sequences for analysis because too few were observed. A summary of the total number of

227

signals coded in each signal category is provided in Table A1.3.

Table A1.3. Total number of signals coded within each signal category across age groups

Age (category)	Signal category				
	Unimodal	Multimodal combination	Fixed multimodal	Unimodal sequence	Multimodal sequence
Infant	472	100	5	98	14
Juvenile	424	117	3	41	18
Early adolescent	189	115	12	48	28
Total	1085	332	20	187	60

Appendix 1.3

Table A1.4. Descriptions of facial expressions

Facial expressions	Description
Open mouth	Lip corners are retracted and parted. Mouth can be open or stretched wide open. In the context of play the eyes and face can be relaxed or tensed depending on play intensity. This may or may not be accompanied by laughter vocalization in the context of play.
Pout	Lips are pushed forwards and parted with the chin raised. Mouth may be slightly open.
Bared teeth (open/closed)	The corners of the mouth are withdrawn, retracting the upper and lower lips to expose both the upper and lower

anterior teeth. The teeth can be parted (open) or not

(closed). Eyes may be open or squinted.

Whimper face*	Lips are funnelled, parted, and partially retracted. Mouth corners are pushed forward, and the mouth is partially open.
Scream face*	Upper lips are raised with corners retracted exposing the upper teeth. The lower lip is depressed exposing the lower teeth. The lips are parted, and the mouth is wide open.
Huu/Pant hoot face*	Lips are pursed forward and parted. The mouth is rounded and can be open or not

Repertoire is based on Parr et al. (2005, 2007) and Bard et al. (2011).

*Facial expressions produced in combination with associated vocal components were coded as fixed multimodal (facial–vocal) signals.

Table A1.5. Descriptions of vocalizations (call types)

Vocalizations (call type)	Description
Bark	Sharp, loud calls with an abrupt onset. They are often noisy and generally low frequency.
Grunt	Low-frequency, short calls which can be produced singularly or in short bouts. Depending on the situation in which they are produced they can vary in tonality, noise and rhythm.
Huu-call	Tonal calls with most energy at onset with a rise and fall in frequency over the call.

Laughter	Low-frequency, noisy grunts and moans produced in series during alternating inhalations and exhalations in an irregular rhythm.
Pant hoot*	A call series with typically four distinct phases: an introductory phase of low frequency hoo calls; a build-up phase with increasingly loud hoo calls with acoustic energy in both inhalation and exhalation; a climax phase including screaming; a let-down phase which resembles the introductory phase but progressively decreasing energy. Introductory and let down phase may be omitted.
Scream*	High-frequency, loud and harmonic vocalization with varying degrees of tonality almost always produced in bouts. Acoustic energy typically only present during exhalation but during intense tantrums it is often present during inhalation as well.
Squeak	High-frequency, short calls predominately given in fast succession to form short bouts. These calls are tonal signals, often with clear harmonic structures.
Whimper*	Low-frequency, soft hoo calls that can vary in frequency and amplitude as a bout progresses. Individual hoo calls are tonal signals with a variable number of harmonics.

Repertoire is based on van Lawick-Goodall (1968), Plooij (1984), Kojima (2008) and Slocombe and Zuberbühler (2010)

*Vocalizations produced in combination with associated facial component were coded as fixed multimodal (facial–vocal) signals.

Table A1.6. Descriptions of observed gestures

Gesture	Description
Arm raise	Arm and or/hand raised vertically in the air
Arm shake	Arm is moved in a small back and forwards motion repeatedly
Arm swing	Arm is moved below the shoulder in a large back and forth motion
Beckon	Hand moved in a sweep from elbow or wrist towards signaller
Big loud scratch	Loud exaggerated scratching movement on signaller's own body
Bipedal stance	Standing bipedally, arms often held out to side with back arched
Bite	Teeth are pressed into the skin of recipient's body
Bow	Signaller bends forward from waist while bipedal
Crouch	Quadrupedal posture, turned towards the recipient with the limbs flexed
Dangle	Signaller hangs from one or both arms from a branch above another individual.
Directed push	A noneffective, light push of percipient which indicates direction of desired movement, followed immediately by the recipient moving as indicated
Drum other	Short hard auditory contact of alternating palms against recipient
Embrace	Both arms are wrapped around a recipient's body and physical contact is maintained
Finger in mouth	Finger(s) is placed into the mouth of the recipient
Gallop	Exaggerated running movement where contact of hands and feet is deliberately auditory

Grab	The hand or foot is closed firmly over a part of the recipient's body (1- or 2-handed)
Grab-pull	As 'Grab' but contact is maintained, and force is exerted to move the recipient from its current position
Hand fling	Hand or arm is moved rapidly in the direction of a recipient
Hand on	Contact of the knuckles or palm of the hand on the body of the recipient for >2 s
Head shake	Repeated back and forth motion of head
Head stand	Body is bent forward with head placed on the ground
Hit with object	An object is brought into short hard contact with the body of the recipient
Jump	Horizontal displacement through the air propelled by both feet
Kick	Hard, short contact of the foot with an object or body of recipient (1 or 2 feet). Can be a forward, sideways or backward movement
Look	Looking intently into the face of a recipient from a few centimetres distance for a minimum duration of 2 s (also described as 'peer')
Mouth stroke	Signaller's palm or fingers repeatedly run over mouth area of recipient
Object in mouth approach	Approaching a recipient while carrying an object in the mouth (e.g., a small branch)
Object move	Object is displaced in one direction. Includes instances where contact is maintained, or item is thrown
Object shake	Repeated back and forth movement of an object (e.g., branch)
Pirouette	Body turns on its vertical axis while also displacing along the ground

Poke	One or more fingers pushed firmly but briefly into the body of the recipient
Pounce	Displacement through the air to land quadrupedally on the body of the recipient
Present body part	Body part is moved to deliberately expose an area to recipient's attention
Present genitals	Recipient is approached backwards, with exposure of the swelling or anus in the direction of the recipient's face
Punch ground	Movement of whole arm, with short hard auditory contact of closed fist to an object or the ground
Push	Forceful contact of the palm on recipient's body in an attempt to displace recipient
Reach	Arm extended in the direction of the recipient with hand opened, palm upwards (no contact)
Roll over	Rolling on to the back exposing the stomach area, often accompanied by repeated movements of the arms and/or legs
Rub rump	Rump area is pushed and/or rubbed with small repeated up and down movements against the body of the recipient
Side roulade	Body is rotated around the head–feet axis while lying on the ground
Slap object/ground	Movement of the arm from the shoulder with hard short contact of the palm(s) to an object or the ground (1- or 2-handed)
Slap other	As 'slap object' but the palm is brought into contact with the recipient's body

Smack lips	Mouth slightly opened and closed rhythmically with an auditory 'smack' sound when the mouth is open
Somersault	Signaller rolls forward in a curled, compact position so the feet are brought above the head returning to a sitting position
Stiff walk	Walk quadrupedally with a slow exaggerated movement a.k.a. swagger
Stomp	The sole of the foot is lifted and brought downward into a short, hard and auditory contact with a surface (2 feet together or alternately)
Stomp other	As Stomp but contact is with recipient's body
Tap other	Single or multiple movements of the arm from the wrist or elbow with short but firm contact on a recipient's body
Touch other	Light contact of the palm and/or fingers on recipient's body for <2 s

Repertoire is based on Nishida et al. (1999) and Byrne et al. (2017).

Table A1.7. The number of individuals that produced different multimodal combination types across contexts (mean signal frequency ± SD)

Multimodal combination	Context			
	Feeding	Play	Aggression	Other
FV	6 (0.21 ± 0.58)	15 (0.77 ± 1.41)	10 (0.31 ± 0.54)	2 (0.15 ± 0.60)
FG	5 (0.19 ± 0.46)	15 (0.76 ± 0.77)	8 (0.20 ± 0.62)	7 (0.59 ± 0.69)
GV	12 (0.76 ± 0.64)	1 (0.02 ± 0.20)	10 (0.64 ± 0.75)	6 (0.41 ± 0.63)
FGV	4 (0.16 ±0.37)	1 (0.02 ± 0.14)	5 (0.20 ± 0.45)	2 (0.11 ± 0.32)

F=facial expression; G=gesture; V=vocalization.

Mean signal frequency is the mean number of each signal type produced in a context across all communicative interactions containing a multimodal combination within that context.
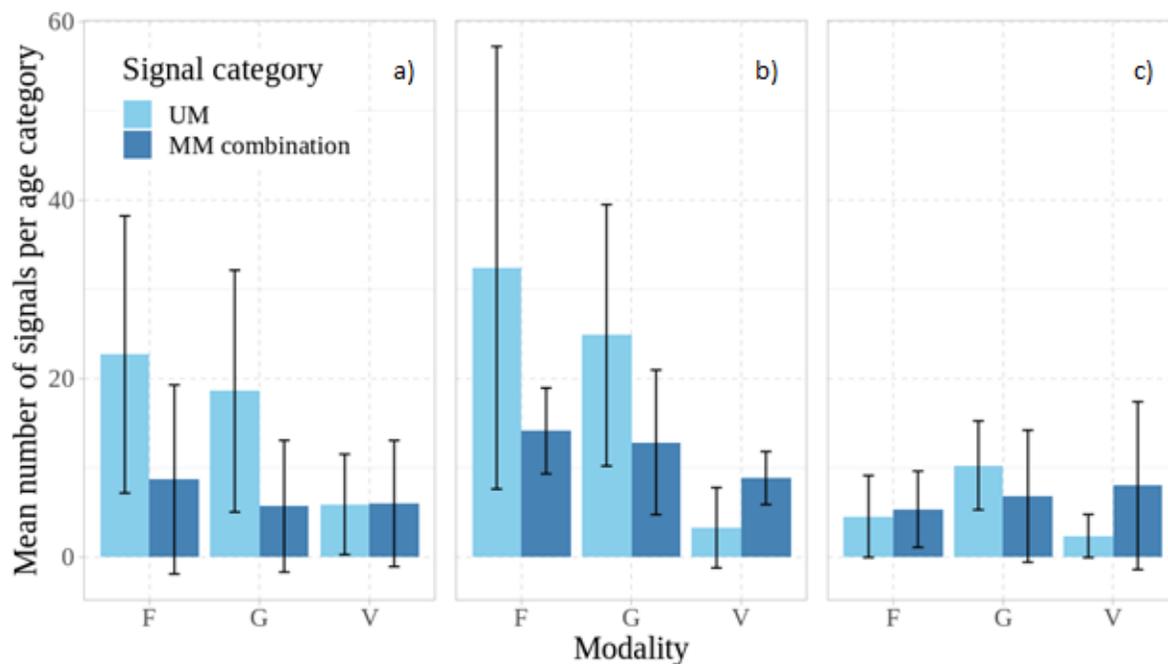
Figure A1.1. The mean frequency per communicative interaction of facial expressions (F), gestures (G) and vocalizations (V) produced as unimodal (UM) signals and as part of multimodal (MM) combinations in (a) infant, (b) juvenile and (c) early adolescent chimpanzees. Error bars represent mean ± 1 SE.

Table A1.8. Factors affecting the frequency of unimodal signals and multimodal combinations across communicative interactions derived using GLMMs with, respectively, a Poisson error structure with a log link function and a zero-inflated, negative binomial distribution.

	Estimate	SE	X^2	P
Unimodal signals				

Intercept	-1.29	0.18	-	-
Age	-0.05	0.05	0.97	0.32
Context [feeding]	0.04	0.12	0.10	0.75
Context [play]	**0.58**	**0.09**	**41.17**	**<0.001**
Context [aggression]	**-0.80**	**0.22**	**17.29**	**<0.001**
Sex [male]	0.03	0.12	0.05	0.82
Group [2]	**-0.33**	**0.13**	**5.63**	**0.02**
Group [4]	**-0.53**	**0.23**	**4.63**	**0.03**
Observation year [2021]	**0.54**	**0.15**	**10.47**	**0.001**
Multimodal combinations				
Intercept	-2.66	0.31	-	-
Age	**0.23**	**0.09**	**6.43**	**0.01**
Context [feeding]	0.14	0.25	0.31	0.58
Context [play]	**0.55**	**0.22**	**7.00**	**0.01**
Context [aggression]	**0.86**	**0.24**	**13.08**	**<0.001**
Sex [male]	**-0.35**	**0.17**	**2.98**	**0.08**
Group [1]	**-0.53**	**-2.55**	**6.14**	**0.013**
Group [4]	-0.48	-1.33	1.77	0.18

Observation year [2021]	0.83	3.42	7.68	**0.01**

Bold values indicate significance
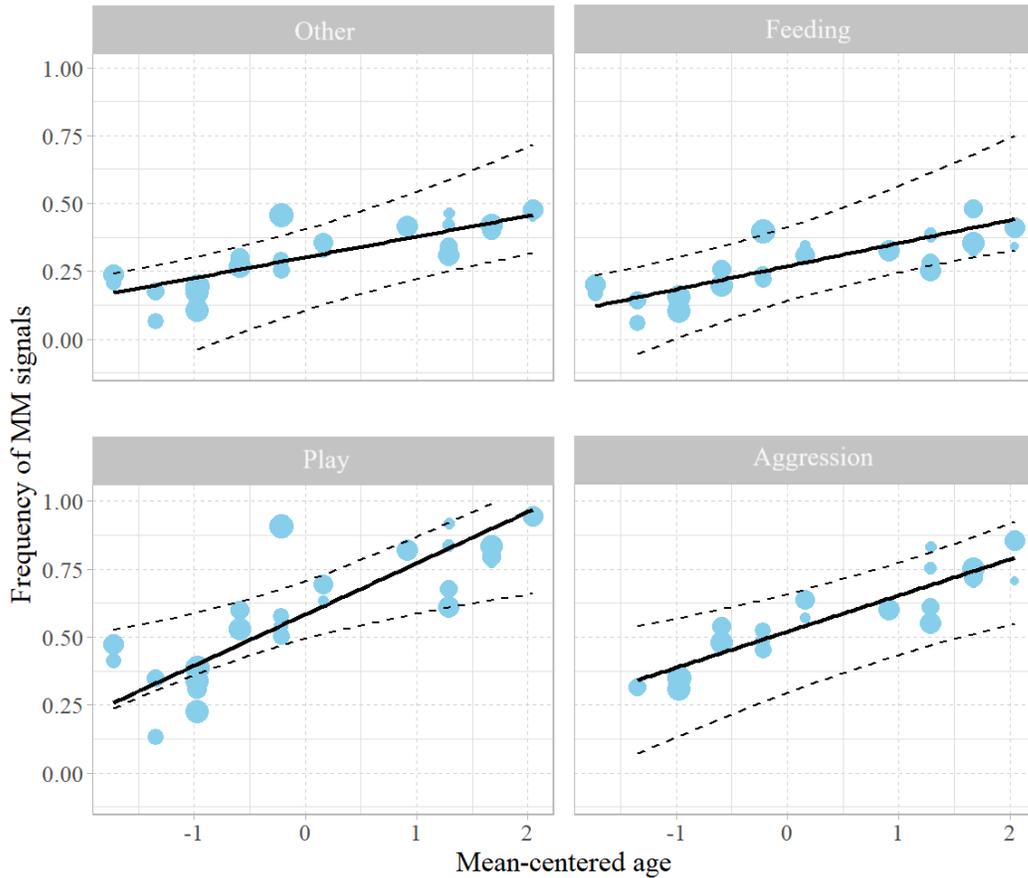
Figure A1.2. Effect of individual age (years) on the frequency of multimodal (MM) combination production in immature chimpanzees ($N = 28$) as a function of context. The mean number of MM combinations produced by each focal individual, in each context is shown. Area of the dots reflects the variation in sample size for each individual for each year of age. The solid line and dashes lines represent the fitted GLMM and 95% confidence intervals, respectively.

Table A1.9. Factors affecting the proportional production of multimodal combinations relative to unimodal signals derived using a GLMM with binomial error structure and a logit link function

	Estimate	**SE**	**X^2**	**P**
Intercept	-1.52	0.29	-	-
Age	0.32	0.11	8.66	**0.003**
Context [feeding]	0.99	0.26	0.14	0.71
Context [play]	0.12	0.23	0.28	0.60
Context [aggression]	1.67	0.34	16.90	**<0.001**
Sex [male]	-0.25	0.23	0.29	0.29
Group [1]	0.17	0.17	0.30	0.30
Group [4]	-0.26	0.24	0.27	0.27
Observation year [2021]	0.18	0.22	0.43	0.43

Bold values indicate significance

Table A1.10. Factors affecting the frequency of unimodal facial expressions and gestures across communicative interactions derived using a GLMM with a Poisson error structure and a log link function

	Estimate	**SE**	**X^2**	**P**
Facial expressions				
Intercept	-3.39	0.36	-	-
Age	-0.10	0.07	1.55	0.21

Context [feeding]	-0.40	0.40	0.99	0.32
Context [play]	2.43	0.27	168.77	**<0.001**
Context [aggression]	0.86	0.43	3.82	0.051
Sex [male]	0.07	0.17	0.68	0.68
Group [1]	-0.46	0.20	0.03	**0.03**
Group [4]	-0.53	0.34	0.14	0.14
Observation year [2021]	0.60	0.23	0.01	**0.01**
Gestures				
Intercept	-1.44	0.19	-	-
Age	0.000	0.06	0.000	0.99
Context [feeding]	-0.16	0.12	1.73	0.19
Context [play]	-0.91	0.13	48.97	**<0.001**
Context [aggression]	-0.57	0.23	6.72	**0.01**
Sex [male]	0.17	0.28	1.60	0.21
Group [1]	-0.11	0.15	0.54	0.46
Group [4]	-0.39	0.23	2.82	0.09
Observation year [2021]	0.58	0.16	11.20	**0.001**

Bold values indicate significance

Diagnostics checks for generalized linear mixed models (GLMM)

Model 1a

Variance inflation factor (VIF) between model variables:

Variable	VIF
Age_year	1.67
Context_feeding	1.48
Context_play	1.43
Context_aggression	1.32
Sex	1.58
Group_1	1.92
Group_4	2.33
Observation_year	2.33

Independence of residuals:

Obtained via 'check_autocorrelation' function in R package 'performance'. Residuals appear to be independent and not autocorrelated (p=0.49).

Distribution of random effect(s):
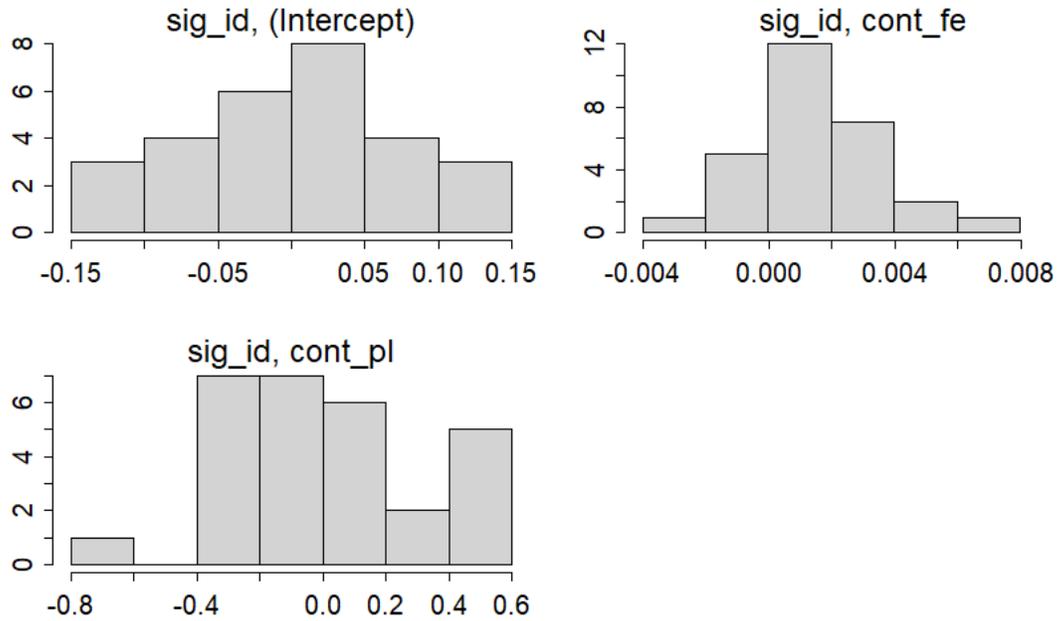
Model 1b

Variance inflation factor (VIF) between model variables:

Variable	VIF
Age_year	1.91
Context_feeding	1.45
Context_play	2.16
Context_aggression	2.16
Sex	1.76
Group_1	1.92
Group_4	2.12
Observation_year	2.28

Independence of residuals:

Obtained via 'check_autocorrelation' function in R package 'performance'. Residuals

appear to be independent and not autocorrelated (p=0.16).

Distribution of random effect(s):

Model 2

Variance inflation factor (VIF) between model variables:

Variable	VIF
Age_year	1.50
Context_feeding	1.38
Context_play	1.83
Context_aggression	1.42
Sex	1.81
Group_1	2.57
Group_4	2.23
Observation_year	3.16

Independence of residuals:

'check_autocorrelation' function in R package 'performance'. Residuals appear to be

independent and not autocorrelated (p=0.09).

Distribution of random effect(s):
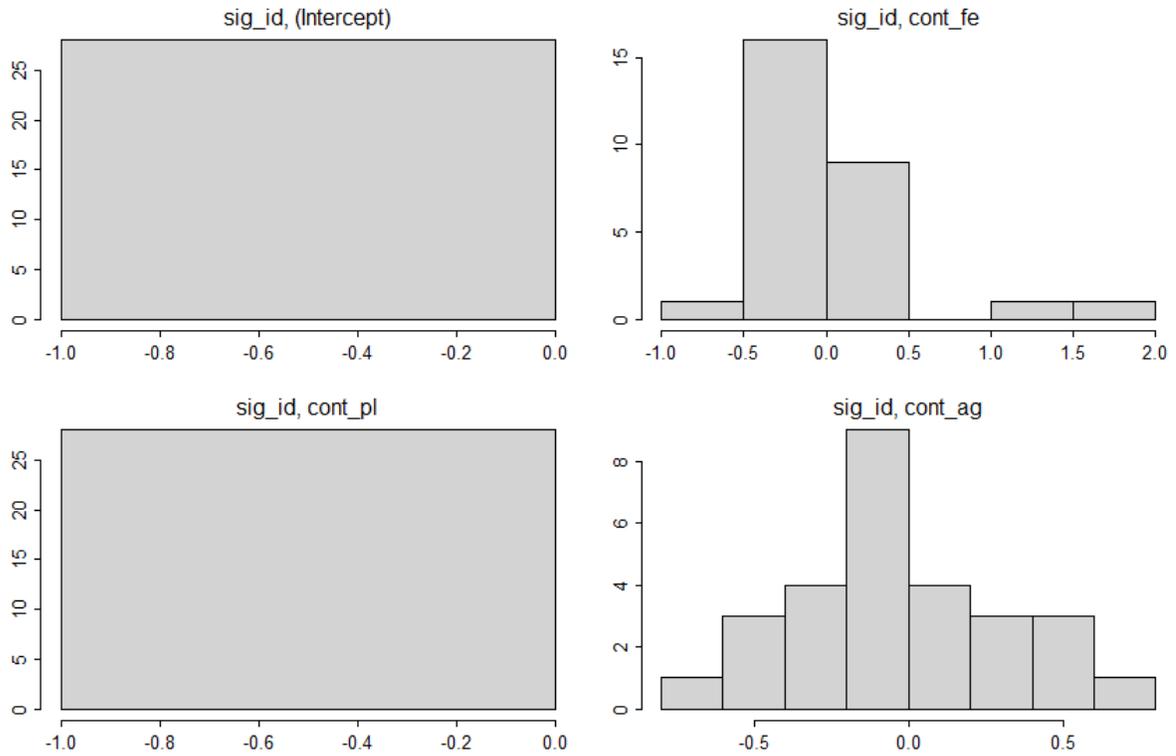
Model 3a

Variance inflation factor (VIF) between model variables:

Variable	VIF
Age_year	1.37
Context_feeding	1.48
Context_play	1.50
Context_aggression	1.01
Sex	1.38
Group_1	1.87
Group_4	2.43
Observation_year	2.31

Independence of residuals:

Obtained via 'check_autocorrelation' function in R package 'performance'. Residuals appear to be independent and not autocorrelated (p=0.82).

Distribution of random effect(s):

sig_id, (Intercept)
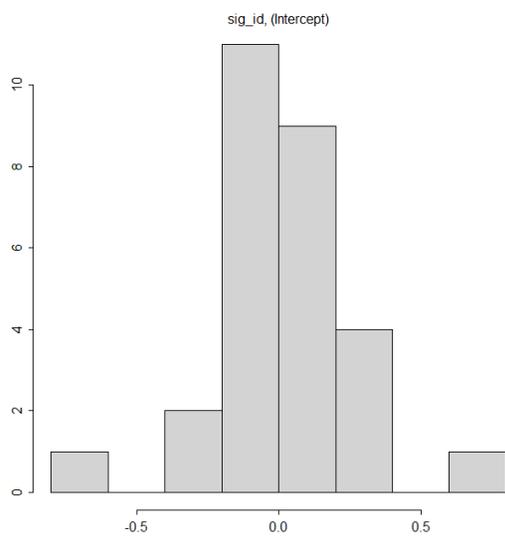
Model 3b

Variance inflation factor (VIF) between model variables:

Variable	VIF
Age_year	1.67
Context_feeding	1.58
Context_play	1.59
Context_aggression	1.27
Sex	1.97
Group_1	2.17

Group_4	2.62
Observation_year	2.47

Independence of residuals:

Obtained via 'check_autocorrelation' function in R package 'performance'. Residuals appear to be independent and not autocorrelated (p=0.08).

Distribution of random effect(s):
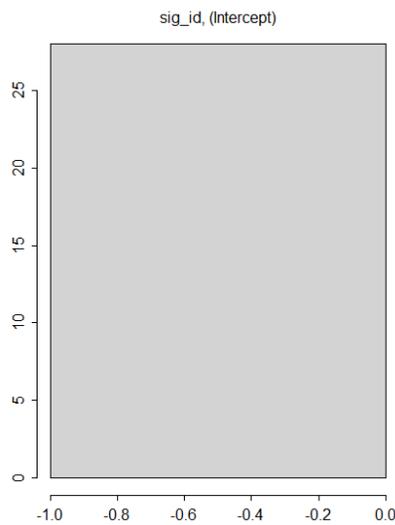
Appendix 2

Appendix 2.1

I intended to test the effect of maternal interactional behaviour in early infancy, on the communicative skills of six infants in later ontogeny for whom we had access to some longitudinal data. I quantified communicative skill using three measures including frequency of engagement in communicative interactions, signal repertoire size and communicative success (i.e., proportion of goal-outcome matches; Cartmill & Byrne, 2010).

I conducted Kendall's tau (τ) correlations. Specifically, I was interested in the relationship between maternal responsiveness at time point 1 (see Table A7) and measures of communication skills at time point 2, including communicative interaction rates, repertoire size and communicative success (i.e., frequency of goal-outcome matches (Catmill & Byrne, 2010).

Table A2.1. Details of individuals included in longitudinal analysis

Infant ID	Mother ID	Age (months) T1	Age (years) T1	Age (years) T2	Observation period (T1/T2)
Chitalu	Claire	27-30	2	6	2017/2021
Don	Debbie	17-20	1	5	2017/2021
Ida	Ingrid	14-17	1	5	2017/2021
Gloria	Genny	15-18	1	5	2017/2021
Kenny	Kambo	24-27	2	6	2013/2017
Tina	Tess	24-27	2	6	2017/2021

I found that individuals with more responsive mothers in early infancy, were more likely to have higher communicative interaction rates as juveniles ($\tau = 0.73$, $P = 0.032$). I found no other significant correlation between maternal responsiveness and repertoire size ($\tau = 0.27$, $P = 0.44$) or communicative success ($\tau = -0.33$, $P = 0.34$). However, the sample size with available data was very small, with as little as 1 case per individual for some matrices of communicative skills.

Appendix 2.2

Table A2.2. Descriptions of 'apparent interactive goals' coded and dyad-members for which they were recorded as marked by 'X'.

Apparent goal	Description	Infant	Mother
Climb on me	Recipient climbs on signallers' body		x
Climb on you	Recipient allows signaller to climb on their body	x	
Move away	Move away from signaller		x
Move closer	Move closer to signaller		x
Contact	Physical affiliative context e.g., when signaller is distressed	x	x
Initiate play	Play with signaller	x	x
Resume play	Previous play bout is resumed with signaller after a pause	x	
Travel with me	Travel together with signaller	x	x
Initiate grooming	Grooming between signaller and recipient	x	x
Groom here	Signaller or recipient focuses on an indicated location		x
Reposition body	Move body into indicated position	x	x
Share food	Feeding recipient shares food item with signaller	x	
Nurse me	Recipient permits nursing	x	
Hang on	Hold body in indicated position		x
Stop that	Cease a current behaviour directed towards signaller	x	x

Goals and descriptions based on Cartmill & Byrne, 2010; Hobaiter & Byrne, 2014.

Table A2.3. Factors affecting the likelihood of chimpanzee mothers using communicative signals instead of physical actions derived using a GLMM's with binomial error structure and a logit link function.

	Estimate	SE	X^2	*P*
Intercept	-1.053	0.522	-	-
Infant age [within]	-0.093	0.138	0.456	0.499
Infant age [between]	0.514	0.190	7.134	**0.008**
Infant sex [Female]	-0.336	0.345	0.997	0.318
Maternal proximity	2.051	0.295	55.821	**<0.001**
Contextual risk	0.398	0.424	0.877	0.349
Mother's parity	0.354	0.132	5.392	**0.020**
Group	0.069	0.349	0.040	0.841
Observation year [2021]	-1.01	0.383	5.767	**0.013**

Bold values indicate significance

Table A2.4. Factors affecting maternal responsiveness in chimpanzees derived using a GLMM's with binomial error structure and a logit link function.

	Estimate	SE	X^2	*P*
Intercept	-1.442	0.698	-	-
Infant age [within]	-0.395	0.159	6.441	**0.011**
Infant age [between]	0.189	0.227	0.697	0.404

Interaction behaviour [signal]	0.839	0.345	5.698	**0.015**
Infant sex [Female]	-0.206	0.379	0.296	0.587
Maternal proximity	1.588	0.382	19.576	**<0.001**
Contextual risk	1.327	0.592	5.691	**0.017**
Mother's parity	-0.075	0.157	0.230	0.632
Group	-0.166	0.336	0.246	0.620
Observation year [2021]	0.442	0.466	0.894	0.344

Bold values indicate significance

Diagnostics checks for generalized linear mixed models (GLMM)

Model 1

Variance inflation factor (VIF) between model variables

Variable	VIF
Age_between	1.34
Age_within	1.02
Context_risk	1.08
Maternal_proximity	1.06
Mother_parity	1.13
Sex	1.20
Group	1.17
Observation_year	1.41

Independence of residuals:

Obtained via 'check_autocorrelation' function in R package 'performance'. Residuals appear to be independent and not autocorrelated (p=0.30).

Distribution of random effect(s):

dyad_id, (Intercept)
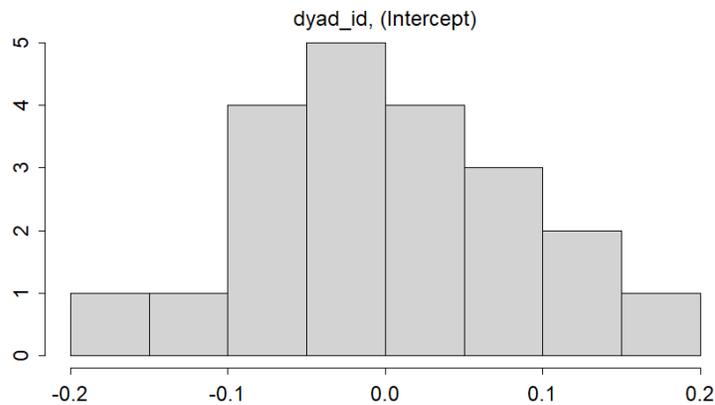
Model 2

Variance inflation factor (VIF) between model variables

Variable	VIF
Age_between	2.20
Age_within	1.13
Interactive_behaviour	1.21
Context_risk	1.09
Maternal_proximity	1.66
Mother_parity	1.18
Sex	1.55
Group	1.24
Observation_year	1.85

Independence of residuals:

Obtained via 'check_autocorrelation' function in R package 'performance'. Residuals appear to be independent and not autocorrelated (p=0.12).

Distribution of random effect(s):

dyad_id, (Intercept)
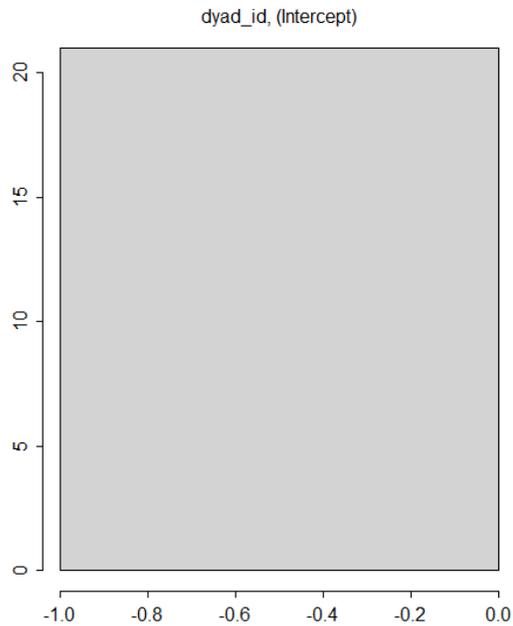
Appendix 3

Table A3.1. Factors affecting multimodal combination frequency in chimpanzees derived using a GLMM with zero-inflated, negative binomial distribution.

	Estimate	**SE**	X^2	P
Intercept	-1.819	0.401	-	-
Age [year]	0.150	0.109	1.813	0.178
Mean interaction rate	-0.173	0.141	1.514	0.219

Number of communication partners	0.267	0.125	4.739	**0.030**
Context [feeding]	0.153	0.272	0.318	0.573
Context [play]	0.833	0.233	14.867	**<0.001**
Context [aggression]	0.945	0.259	13.157	**<0.001**
Sex [female]	0.250	0.196	1.625	0.202
Group [1]	-0.162	0.245	0.438	0.508
Group [4]	0.601	0.413	2.105	0.147
Observation year [2021]	0.110	0.294	0.318	0.710

Bold values indicate significance

Table A3.2. Factors affecting multimodal combination repertoire size in chimpanzees derived using a GLM with a Poisson error structure and a logit-link function.

	Estimate	**SE**	**t value**	*P*
Intercept	-0.698	0.567	-	-
Age [year]	0.028	0.036	0.785	0.442
Mean interaction rate	0.040	0.026	1.563	0.134
Number of communication partners	0.081	0.033	2.498	**0.021**
Sex [female]	-0.026	0.214	-0.120	0.906
Group [1]	-0.210	0.255	-0.822	0.421
Group [4]	0.625	0.449	1.392	0.179
Observation year [2021]	0.308	0.335	0.918	0.370

Bold values indicate significance

Appendix 4

Table A4.1. Description of recipient response types

Response type	Description
Signal	Signal responses corresponded to behavioural definitions of facial expressions, gestures, vocalizations and multimodal combinations applied to focal subjects (see Appendix 1, Table A4-6)
Positive	Positive responses include affiliative behaviours, grooming and play. Responses were also coded as positive if they corresponded to the signaller's communicative goal (see Table A2) e.g., share food.
Negative	Negative responses include aggressive (e.g., threats, physical aggression) or submissive behaviours (e.g., running away, screaming).

Response categories adapted from Wilke et al., 2017.

Table A4.2. Description of apparent communicative goals coded

Apparent goal	Description
Climb on me	Recipient climbs on signallers' body
Climb on you	Recipient allows signaller to climb on their body
Move away	Move away from signaller
Move closer	Move closer to signaller

Contact	Physical affiliative context e.g., when signaller is distressed		
Initiate play	Play with signaller		
Resume play	Previous play bout is resumed with signaller after a pause		
Travel with me	Travel together with signaller		
Initiate grooming	Grooming between signaller and recipient		
Groom here	Signaller or recipient focuses on an indicated location		
Reposition body	Move body into indicated position		
Share food	Feeding recipient shares food item with signaller		
Sexual attention	Recipient responds sexually e.g., copulation		
Nurse me	Recipient permits nursing		
Stop that	Cease a current behaviour directed towards signaller		
Submission	Recipient does not respond with aggression		

Goals and descriptions based on Cartmill & Byrne, 2010; Hobaiter & Byrne, 2014.

Table A4.3. Factors affecting the production of multimodal combinations in chimpanzee signal events derived using GLMM with Binomial error structure and a logit link function.

	Estimate	SE	X^2	P
Intercept	-2.009	0.563	-	-
Age	0.571	0.225	6.651	**0.010**
Recipient visually oriented [yes]	0.781	0.375	4.914	**0.027**
Kin [mother]	-1.110	0.438	0.062	0.803
Kin [maternal kin]	0.121	0.450	0.072	0.789
Context [feeding]	-0.694	0.389	3.623	0.071

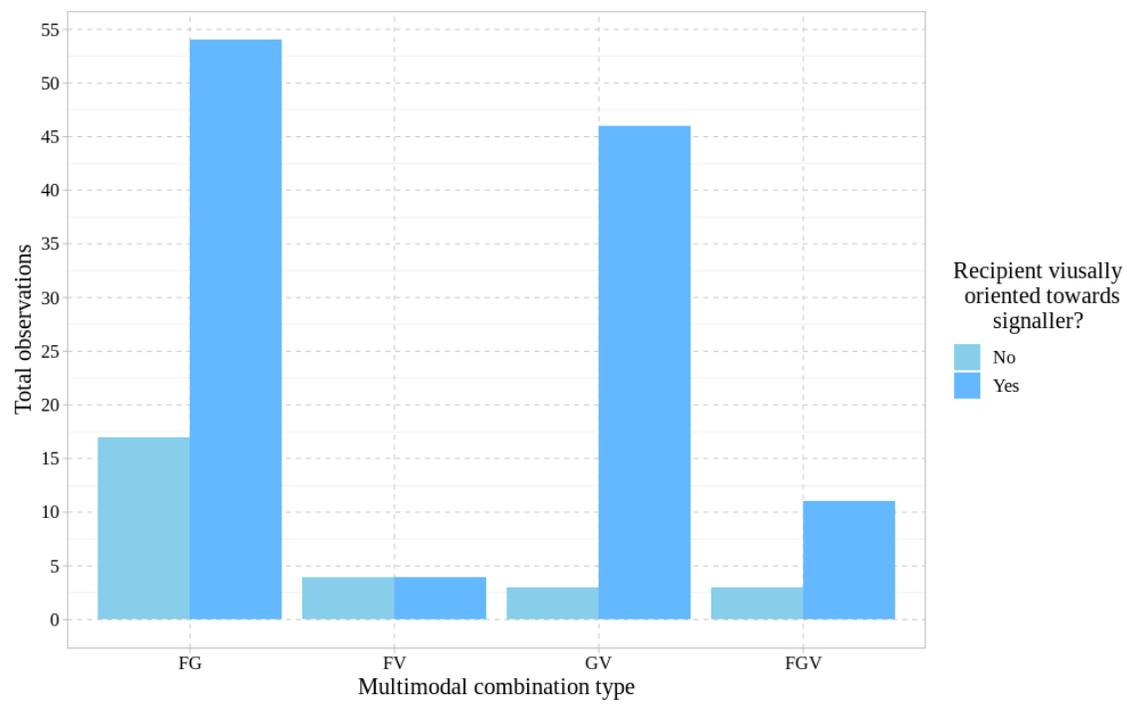|				
Context [play]	1.186	0.390	8.316	**0.004**
Sex [female]	-0.363	0.399	0.836	0.361
Group [1]	-0.035	0.510	0.005	0.946
Group [4]	0.449	0.799	0.315	0.575
Observation year [2017]	-0.283	0.563	0.245	0.620

Bold values indicate significance.

Figure A4.1. Total number of multimodal combinations observed when recipient was visually oriented towards or away from a signaller. F = facial expression, G = gesture, V = vocalization.

Table A4.4. Factors affecting recipient responses to chimpanzee signal events derived using GLMM with Binomial error structure and a logit link function.

	Estimate	SE	X^2	P
Intercept	-1.641	0.383	-636	-
Age	0.147	0.164	0.794	0.373
Multimodal combination	0.568	0.260	4.893	**0.027**
Recipient visually oriented [yes]	0.826	0.280	9.224	**0.002**
Kin [mother]	1.282	0.397	11.725	**<0.001**
Kin [maternal kin]	0.187	0.377	0.247	0.620
Context [feeding]	-0.615	0.426	4.412	**0.036**
Context [play]	1.437	0.302	19.671	**<0.001**
Sex [female]	0.924	0.324	6.807	**0.009**
Group [1]	0.281	0.426	0.446	0.504
Group [4]	1.400	0.626	4.337	**0.037**
Observation year [2017]	-0.569	0.400	1.568	0.211

Bold values indicate significance.

Table A4.5. Factors affecting recipient response latency (ms) to chimpanzee signal events derived using GLMM with Poisson error structure and a log link function.

	Estimate	SE	X^2	P
Intercept	6.70	0.328	-	-

	Estimate	SE	X^2	P
Age * Multimodal combination	-0.200	0.027	55.019	**<0.001**
Recipient visually oriented [yes]	-0.018	0.157	0.014	0.907
Kin [mother]	0.094	0.302	0.095	0.758
Kin [maternal kin]	-0.048	0.332	0.020	0.887
Context [feeding]	0.497	0.257	3.549	0.059
Context [play]	-0.055	0.305	0.032	0.858
Sex [female]	-1.755	0.389	0.207	0.649
Group [1]	-0.002	0.405	0.000	0.995
Group [4]	-0.768	0.608	1.603	0.205
Observation year [2017]	0.551	0.420	1.748	0.186

Bold values indicate significance.

Table A4.6. Factors affecting the achievement of communicative goals derived using GLMM with Binomial error structure and a logit link function.

	Estimate	SE	X^2	P
Intercept	-0.147	0.643	-	-
Age	0.215	0.242	0.791	0.374
Multimodal combination	0.868	0.360	6.214	**0.013**
Recipient visually oriented [yes]	0.921	0.401	5.809	**0.015**
Kin [mother]	1.423	0.498	8.020	**0.005**

Kin [maternal kin]	0.924	0.589	2.517	0.113
Context [feeding]	-2.267	0.543	23.002	**<0.001**
Context [play]	-0.088	0.450	0.038	0.845
Sex [female]	0.322	0.503	0.394	0.530
Group [1]	0.171	0.554	0.096	0.758
Group [4]	0.140	0.896	0.024	0.877
Observation year [2017]	-0.449	0.581	0.589	0.423

Bold values indicate significance.

**Diagnostics checks for generalized linear mixed models (GLMM)**

Model 1

Variance inflation factor (VIF) between model variables

Variable	VIF
Age_year	2.02
Recipient_view	1.13
Kin_mother	1.27
Kin_maternal_kin	1.15
Sex	1.66
Context_feeding	1.43
Context_play	1.58
Group_1	2.52
Group_4	2.75
Observation_year	2.09

Independence of residuals:

Obtained via 'check_autocorrelation' function in R package 'performance'. Residuals

appear to be independent and not autocorrelated (p=0.09).

Distribution of random effect(s):
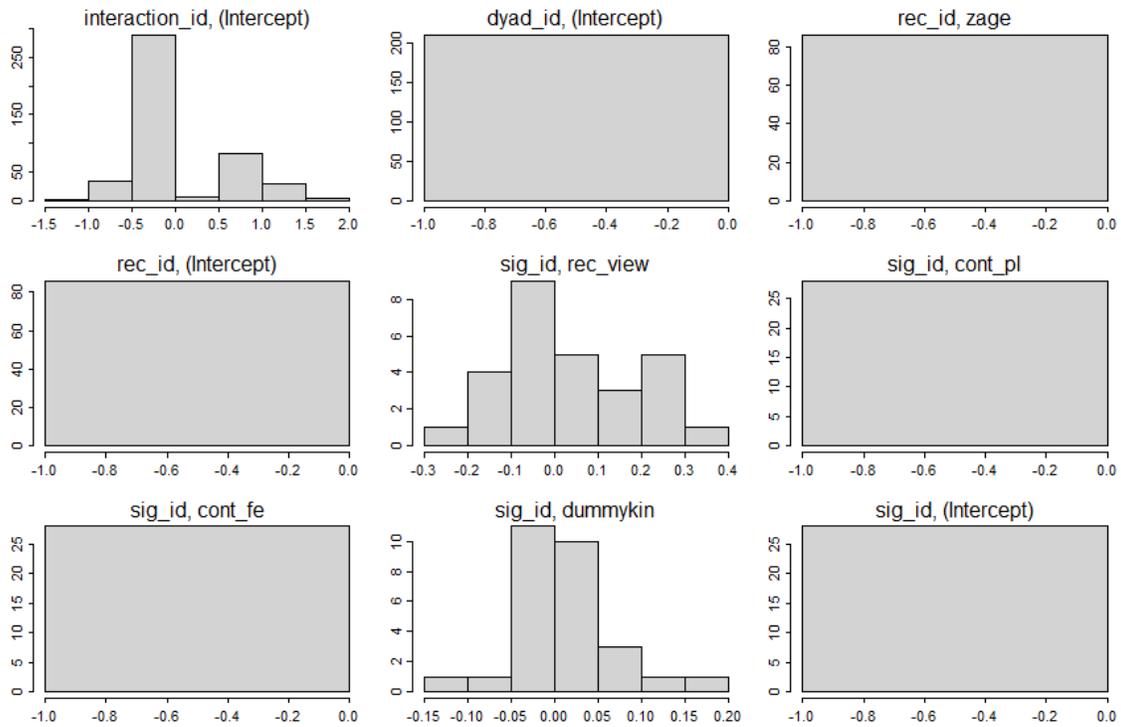
Model 2

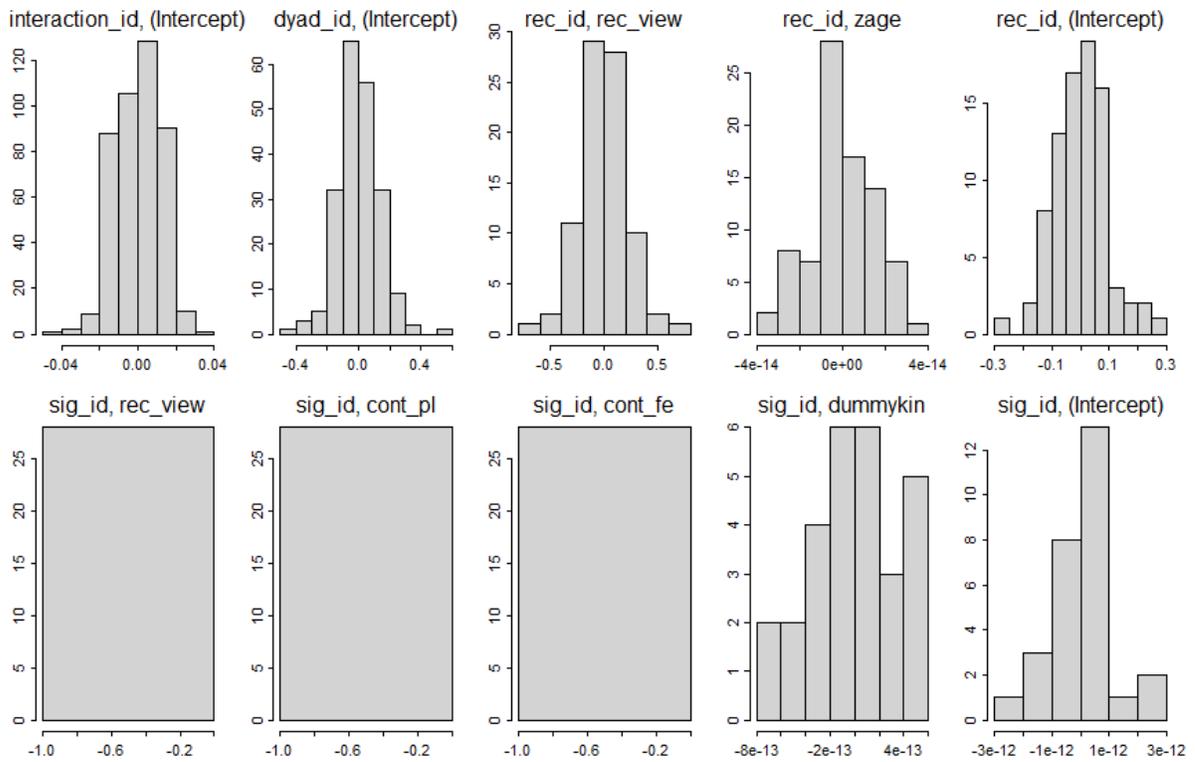*Variance inflation factor (VIF) between model variables*

Variable	VIF
Age_year	1.88
Multimodal	1.08
Recipient_view	1.11
Kin_mother	1.38
Kin_maternal_kin	1.12

Sex	1.64
Context_feeding	1.58
Context_play	1.30
Group_1	2.38
Group_4	3.13
Observation_year	2.58

Independence of residuals:

Obtained via 'check_autocorrelation' function in R package 'performance'. Residuals appear to be independent and not autocorrelated (p=0.56).
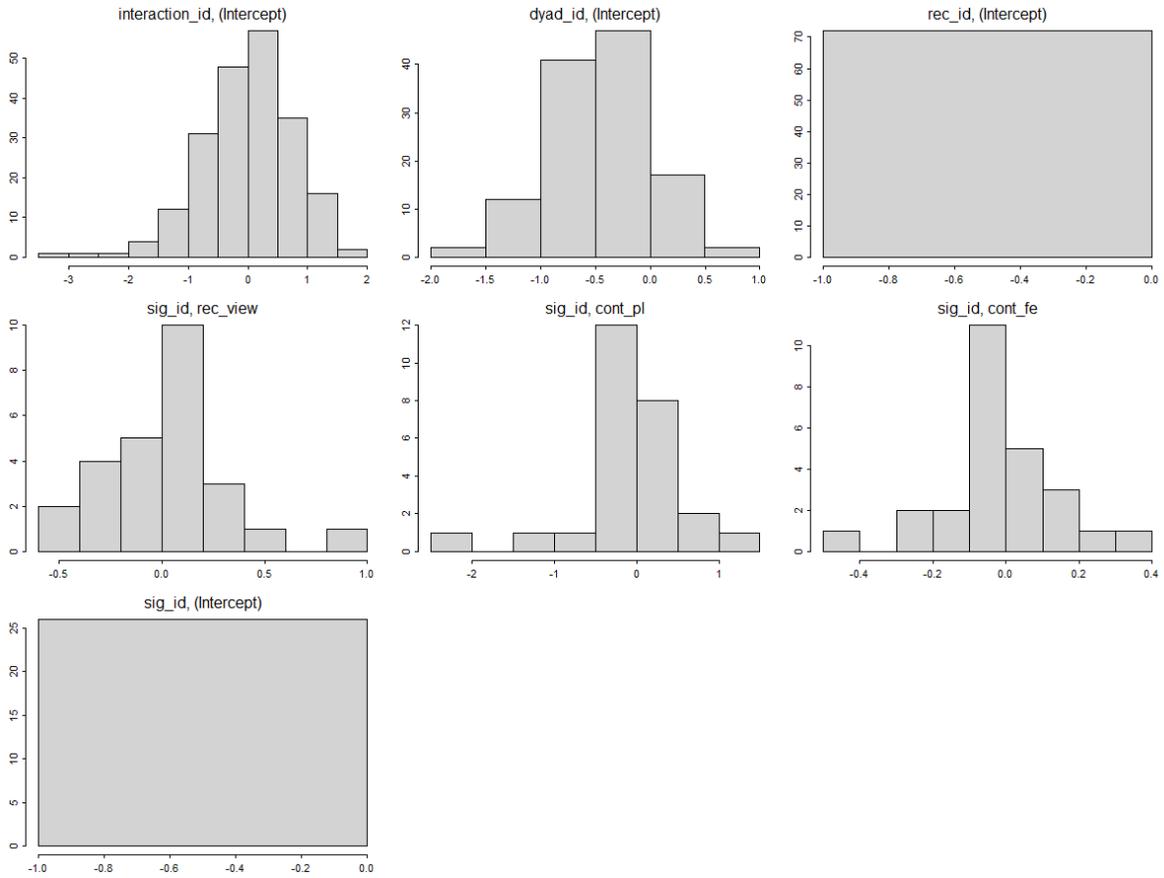
Distribution of random effect(s):

Model 3

Variance inflation factor (VIF) between model variables

Variable	VIF
Age:Multimodal	1.24
Recipient_view	1.17
Kin_mother	1.18
Kin_maternal_kin	1.14
Sex	2.05
Context_feeding	1.21
Context_play	1.10
Group_1	2.46
Group_4	3.15
Observation_year	2.71

Independence of residuals:

Obtained via 'check_autocorrelation' function in R package 'performance'. Residuals appear to be independent and not autocorrelated (p=0.08).

Distribution of random effect(s):

Model 4

Variance inflation factor (VIF) between model variables

Variable	VIF
Age_year	1.66
Multimodal	1.07
Recipient_view	1.16
Kin_mother	1.63
Kin_maternal_kin	1.30
Sex	1.92
Context_feeding	2.00

Context_play	1.58
Group_1	2.32
Group_4	2.74
Observation_year	2.56

Independence of residuals:

Obtained via 'check_autocorrelation' function in R package 'performance'. Residuals appear to be independent and not autocorrelated (p=0.15).

Distribution of random effect(s):