Animal Behaviour 201 (2023) 175–190

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Multimodal communication development in semiwild chimpanzees

Emma Doherty [a], Marina Davila-Ross [b], Zanna Clay [a, *]

[a] Department of Psychology, Durham University, Durham, U.K.
[b] Psychology Department, University of Portsmouth, Portsmouth, U.K.

ARTICLE INFO

Article history:
Received 30 June 2022
Initial acceptance 27 September 2022
Final acceptance 8 February 2023
Available online 5 June 2023
MS. number: 22-00327R

Keywords:
great ape
language evolution
multimodal communication
ontogeny
signal production

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.

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 our 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 our 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. Crockford et al., 2015; Fitch, 2017; Levinson & Holler, 2014; Townsend et al., 2017, 2020; Van Schaik, 2016) 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 we 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 Frohlich & 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

* Corresponding author.
E-mail address: zanna.e.clay@durham.ac.uk (Z. Clay).

https://doi.org/10.1016/j.anbehav.2023.03.020

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 (Partan & Marler, 2005; Smith, 1977) 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; Oña et al., 2019; Wilke et al., 2017; Fröhlich et al., 2019; Oña et al., 2019), we consider multimodal signal combinations (henceforth 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 our 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 et al., 2015; Frohlich et al., 2019; 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 our 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 our closest primate relatives can have important implications for our 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; Fröhlich et al., 2016) 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. Acredolo & Goodwyn, 1988; Bretherton & Bates, 1979; Esteve-Gibert & Prieto, 2014; Iverson & Goldin-Meadow, 2005; Iverson & Thelen, 1999; 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 (2011) 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., 2016). 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, 2015; Hobaiter et al., 2017; Wilke et al., 2017; Genty, 2019; Oña et al., 2019). If this is the case, we could expect 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 (Pusey, 1990; van Lawick-Goodall, 1968) 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 (Bard et al., 2011; Parr et al., 2005, 2007), gestures (Byrne et al., 2017; Nishida et al., 1999) 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, with approval (see Ethical Note), for noninvasive 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 short 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 our 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 our 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 our 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 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).

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, Table A1 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 one in which a signal was produced. 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 (Hobaiter et al., 2017; Schneider et al., 2012a, 2012b), we coded nine behavioural contexts (Table 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, we 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 our analyses. Remaining contexts were included in an 'other' context category which was used as the reference level in our statistical models. See Appendix 1, Table A2 for full reporting of the 'other' category.

Table 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

Table 2
Description of behavioural contexts for communication

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 threats or physical agonistic encounters. Includes submissive signals
Feeding[c]	Individuals engaged in behaviours related to food intake (e.g. begging behaviour). Includes also nursing-related behaviour
Grooming[a,b,c]	Behaviour accompanying the request of or participation in grooming interactions
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 including play signalling, noncontact (e.g. chasing) or contact (e.g. wrestling) play
Sexual[a,b]	Behaviour accompanying sexual interaction, e.g. presenting genitals
Travel[a,b]	Behaviour accompanying locomotion in the enclosure. Excludes movement associated with other contexts (e.g. play chasing or running away from an aggressor)

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

Behavioural Coding

A total of 950 clearly visible communicative interactions were observed across 390 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, 2011; 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 2 including definitions and justifications for exclusion. For information on the total number of signals coded within each signal category see Appendix 2, Table A3.

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 3. Full descriptions of all signals observed are given in Appendix 3, Tables A4, A5 and A6. Six types of facial expressions were used in this analysis, based upon prototypical chimpanzee expressions (Parr et al., 2005) 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 3, Table A5). 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 52 gestures were coded from our video footage based largely on the repertoire proposed by Byrne et al., 2017 (see Appendix 3, Table A6). For details of fixed multimodal and different multimodal combinations coded including the context in which they were produced, see Table 4.

Intercoder 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 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 et al., 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

Table 3
Signal types produced within each modality

Signal modality	Total	Signal types
Facial expressions	6	Bared teeth face, open mouth face, pant hoot face, pout face, scream face, whimper face
Vocalizations	8	Bark, grunt, huu-call, laughter, pant hoot, squeak, scream, whimper
Gestures	52	Arm raise, arm shake, arm swing, beckon, big loud scratch, bipedal rocking, bipedal stance, bite, crouch, dangle, directed push, drum other, embrace, finger in mouth, gallop, grab, grab-pull, hand fling, hand on, head shake, head stand, hip thrust, hit with object, jump, kick, look, mouth stroke, object in mouth, 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, stiff walk, stomp, stomp other, tandem walk , tap other, throw object, touch other

Table 4
Repertoire of multimodal (MM) combinations and fixed MM signals produced by chimpanzees aged 1–11 years at Chimfunshi Wildlife Orphanage

Age	Signal combination	Total no. of individuals	Context			
			Feeding	Play	Aggression	Other
MM combination						
Infant (1–4 years, $N = 10$)	FV					
	Open mouth + laugh	40 (7)		X		
	Bared teeth + laugh	1 (1)		X		
	FG					
	Open mouth + bipedal rocking	2 (1)		X		
	Open mouth + grab	1 (1)		X		
	Open mouth + hand fling	1 (1)		X		
	Open mouth + head stand	1 (1)		X		
	Open mouth + head shake	1 (1)		X		
	Open mouth + hit with object	1 (1)		X		
	Open mouth + object move	1 (1)		X		
	Open mouth + pounce	2 (2)		X		
	Open mouth + roll over	3 (2)		X		
	Open mouth + slap other	6 (4)		X		
	Open mouth + somersault	4 (3)		X		
	Open mouth + stomp	2 (1)		X		
	Pout + directed push	1 (1)				X
	Pout + hand fling	1 (1)				X
	Pout + hand on	2 (1)				X
	Pout + poke	1 (1)				X
	Pout + slap other	1 (1)			X	
	Pout + touch other	1 (1)	X			
	GV					
	Directed push + whimper[a]	1 (1)	X			
	Touch other + whimper	2 (1)	X			
	Reach + whimper	1 (1)				X
	Slap other + grunt	1 (1)	X			
	Bite + grunt	1 (1)	X			
	Slap other + scream[a]	1 (1)	X			
	Grab + scream[a]	1 (1)			X	
	FGV					
	Pout + stomp + bark	1 (1)			X	
	Bared teeth + touch other + squeak	1 (1)			X	
	Bared teeth + reach + whimper	1 (1)	X			
Juveniles (5–7 years, $N = 7$)	FV					
	Bared teeth + squeak	6 (4)	X		X	
	Open mouth + grunt	1 (1)			X	
	Open mouth + laugh	16 (3)		X		
	Pout + grunt	1 (1)		X		
	FG					
	Bared teeth + hand fling	1 (1)			X	
	Open mouth + arm raise	2 (2)		X		
	Open mouth + arm shake	1 (1)		X		
	Open mouth + arm swing	1 (1)		X		
	Open mouth + dangle	1 (1)		X		
	Open mouth + drum other	1 (1)			X	
	Open mouth + grab	2 (2)		X		
	Open mouth + grab pull	1 (1)		X		
	Open mouth + object move	1 (1)		X		
	Open mouth + pirouette	1 (1)		X		
	Open mouth + roll over	5 (4)		X		
	Open mouth + slap other	5 (4)		X		
	Open mouth + somersault	12 (4)		X		
	Open mouth + stiff walk	1 (1)		X		
	Open mouth + stomp	6 (3)		X		
	Open mouth + tap other	4 (3)		X		
	Open mouth + touch other	2 (2)		X		
	Pout + reach	1 (1)	X			
	Pout + slap other	1 (1)		X		
	Pout + tap other	1 (1)				X
	Pout + touch other	4 (3)		X		X
	Reach + squeak	1 (1)	X			
	Tap other + squeak	1 (1)			X	
	GV					
	Arm raise + grunt	1 (1)				x
	Arm raise + whimper[a]	1 (1)				
	Bipedal rock + whimper	1 (1)				X
	Bite + squeak	1 (1)				X
	Drum other + grunt	1 (1)	X			
	Reach + grunt	1 (1)	X			
	Reach + squeak	1 (1)	X			

(continued on next page)

Table 4 (*continued*)

Age	Signal combination	Total no. of individuals	Context			
			Feeding	Play	Aggression	Other
	Reach + whimper	6 (3)		X		
	Slap ground + pant hoot[a]	1 (1)				X
	Slap ground + scream[a]	1 (1)			X	
	Tap other + squeak	3 (1)				
	Touch other + grunt	3 (3)			X	X
	Touch other + whimper[a]	5 (1)	X		X	
	FGV					
	Bared teeth + arm raise + squeak	1 (1)	X			
	Bared teeth + embrace + squeak	1 (1)			X	
	Bared teeth + reach + whimper	1 (1)	X			
	Pout + reach + grunt	1 (1)				X
	Pout + slap ground + grunt	1 (1)	X			
	Pout + tap other + grunt	1 (1)				X
Early adolescents (8–11 years, $N = 11$)	FV					
	Bared teeth + bark	1 (1)			X	
	Bared teeth + grunt	1 (1)	X			
	Bared teeth + squeak	16 (8)	X		X	X
	Open mouth + laugh	18 (3)		X		
	Open mouth + squeak	1 (1)			X	
	Pout + grunt	1 (1)				X
	FG					
	Bared teeth + present genitals	1 (1)				X
	Bared teeth + slap other	1 (1)			X	
	Open mouth + drum other	1 (1)		X		
	Open mouth + grab pull	1 (1)		X		
	Open mouth + hand fling	1 (1)			X	
	Open mouth + hit with object	2 (1)			X	
	Open mouth + poke	1 (1)		X		
	Open mouth + roll over	1 (1)		X		
	Open mouth + slap other	3 (2)			X	X
	Pout + arm swing	1 (1)				X
	Pout + hand fling	1 (1)				X
	Pout + object move	1 (1)				X
	Pout + reach	2 (2)	X			X
	Pout + rub rump	1 (1)				X
	Pout + slap object	1 (1)			X	
	Pout + slap other	1 (1)		X		
	Pout + stomp	3 (2)			X	X
	GV					
	Bite + squeak	2 (1)			X	
	Crouch + bark	4 (2)			X	
	Crouch + grunt	5 (3)			X	
	Crouch + pant hoot[a]	1 (1)			X	
	Crouch + tap other + grunt	1 (1)			X	
	Directed push + whimper	1 (1)				X
	Hand fling + grunt	1 (1)			X	
	Hand fling + scream[a]	1 (1)			X	
	Hand on + scream	2 (2)			X	
	Hand on + squeak	1 (1)			X	
	Object move + grunt	1 (1)			X	
	Present genitals + squeak	1 (1)	X			
	Reach + grunt	1 (1)	X			
	Reach + scream[a]	1 (1)			X	
	Reach + whimper[a]	1 (1)	X			
	Reach + whimper[a]	1 (1)	X			
	Slap ground + scream[a]	1 (1)			X	
	Slap other + grunt	1 (1)		X		
	Stomp + grunt	1 (1)			X	
	Tap other + bark	1 (1)			X	
	Tap other + grunt	1 (1)				X
	Touch other + bark	2 (2)			X	X
	Touch other + grunt	1 (1)				X
	Touch other + scream[a]	1 (1)	X			
	Touch other + whimper	1 (1)	X			
	Touch other + whimper	2 (2)	X			
	FGV					
	Open mouth + somersault + laugh	1 (1)		X		
	Open mouth + grab pull+ laugh	1 (1)		X		
	Bared teeth + grab + squeak	1 (1)			X	
	Bared teeth + hand fling + squeak	2 (1)			X	
	Bared teeth + reach + whimper	1 (1)	X			
	Bared teeth + touch other + whimper	1 (1)	X			
	Bared teeth + tap other + squeak	3 (2)			X	
	Bared teeth+ touch other + squeak	1 (1)			X	

Table 4 (*continued*)

Age	Signal combination	Total no. of individuals	Context			
			Feeding	Play	Aggression	Other
	Pout + crouch + grunt	1 (1)			X	
	Pout + touch other + grunt	1 (1)				X
Fixed FV signals						
	Pant hoot face + pant hoot	2 (1)			X	
	Scream face + scream	5 (2)	X		X	
	Whimper face + whimper	6 (4)	X			

F = Facial expression; G = gesture; V = vocalization.
[a] Include instances of fixed MM F + V components but included as (free) MM combination when combined with a gesture.

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 (Barr et al., 2013; Schielzeth & Forstmeier, 2009). 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 from the full model 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 0 s 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 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 3, Table A7). 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 (Schielzeth, 2010). We assessed the collinearity between our predictor variables by determining variance inflation factors (VIF, Quinn & Keough, 2002) from a model including only these variables using the function vif of the R package 'car'. Collinearity was not an issue (maximum VIF = 1.69). 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 by comparing estimates obtained from each model based on all data with respective models with the level of random effects excluded one at a time. This revealed no serious stability issues across our models. All models were run in R v. 2.15 (The R Foundation for Statistical Computing, Vienna, Austria, http://www.r-project.org).

Ethical Note

This research was purely observational and did not intervene in chimpanzees' daily activities. All research complied with the Chimfunshi Wildlife Orphanage regulations and with ethical approval from the Animal Welfare Ethical Review Board at Durham University and Chimfunshi Research Advisory Board.

RESULTS

Overview of Signals

Unimodal signals

In total, we observed 1058 unimodal signals across all individuals and contexts. Unimodal signals were the most produced signal category across all ages (*N*, individual mean ± SD: infants: 437, 43.70 ± 30.40; juveniles: 428, 61.14 ± 36.45; early adolescents: 195, 17.27 ± 6.89). Across all age groups, facial expressions were the most common unimodal signal recorded (505, 18.03 ± 18.09) followed by gestures (450, 16.07 ± 12.95) and vocalizations (103, 3.67 ± 3.90). Across contexts, the highest number of unimodal signals were produced in the play context in all age categories (Table 5).

Multimodal combinations

We observed a total of 316 multimodal combinations, making them rare relative to unimodal signals (see Appendix 3, Fig. A1). However, multimodal combinations occurred in 26 of 28 individuals across all age categories (*N*, individual mean ± SD: infants: 86, 8.60 ± 8.38; juveniles: 113, 16.00 ± 10.73; early adolescents: 117, 10.63 ± 10.8.32). We recorded a total of 101 different combinations which included up to six different facial

Table 5
Individual mean number (SD) of unimodal signal modalities produced per communicative interaction across age categories and contexts

Age category (*N*)	Modality	Context			
		Feeding	Play	Aggression	Other
Infant (10)	F	0.10 (0.32)	21.60 (16.04)	0.10 (0.32)	0.10 (0.32)
	G	5.80 (4.05)	7.20 (6.37)	0.33 (0.50)	3.00 (3.68)
	V	0.70 (1.34)	4.40 (4.81)	0	0.40 (0.97)
Juvenile (7)	F	1.29 (1.38)	29.57 (2.40)	0.43 (0.79)	1.29 (2.98)
	G	10.43 (7.68)	7.29 (6.37)	0.43 (0.79)	8.00 (8.52)
	V	0.29 (0.76)	0.86 (1.46)	0.29 (0.76)	0.86 (1.86)
Early adolescent (11)	F	0.09 (0.30)	4.27 (4.86)	0.55 (1.04)	0.36 (0.67)
	G	3.00 (4.73)	1.27 (1.10)	1.64 (2.25)	3.45 (2.38)
	V	0.36 (0.92)	1.18 (2.75)	1.18 (1.66)	0.18 (0.40)

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

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 (111, 3.96 ± 4.37) and facial–gestural–vocal signals the least (22, 0.79 ± 1.26). Infants and juveniles produced the highest number of multimodal combinations in the play context, but early adolescents produced the most in aggressive contexts (Table 6).

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 950 data points corresponding to the total number of communicative interactions observed across 28 individuals. Overall, the full models explained a significant amount of variation in the frequency of unimodal and multimodal combination production (LRT comparing the full and null model for unimodal signal frequency: UM: $\chi^2_7 = 101.34$, $P < 0.001$; MM combination: $\chi^2_7 = 26.22$, $P < 0.01$). In neither model was the interaction term between age and context significant. In terms of signaller age, we found no significant effect on the frequency of unimodal signal production (estimate ± SE = −0.05 ± 0.05, $\chi^2_{12} = 0.97$, $P = 0.32$); age, however, had a significantly positive effect on the frequency of multimodal combination production (0.23 ± 0.09; $\chi^2_{12} = 6.43$, $P = 0.011$; Fig. 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 (0.58 ± 0.09; $\chi^2_{12} = 41.17$, $P < 0.001$) and significantly less in the aggression context (−0.80 ± 0.22; $\chi^2_{12} = 17.29$, $P < 0.001$) than the reference category (Fig. 2a). Multimodal combinations were also produced significantly more in the play context than the reference category (0.55 ± 0.22; $\chi^2_{12} = 7.00$, $P = 0.01$) but, in contrast to unimodal signals, were produced at significantly higher frequencies in the aggressive context (0.86 ± 0.24; $\chi^2_{12} = 13.08$, $P < 0.001$; Fig. 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 3, Table A8.

Production of MM Combinations Relative to UM Signals

We constructed a model to test whether the proportional production ofmultimodal 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 1375 data points representing all instances of unimodal signals and multimodal combinations produced across communicative interactions. Overall, the full model explained significantly more variation in the response than the null model ($\chi^2_7 = 28.81$, $P < 0.001$); here again, however, we found no significant interaction between age and context, so the interaction term was removed from the full model.

Signaller age had a significant positive effect on proportional production of multimodal combinations compared to unimodal signals (0.32 ± 0.11; $\chi^2_{12} = 8.66$, $P = 0.003$; Fig. 3). Again, while we found no significant interaction between age and context, the proportion of multimodal combinations produced relative to unimodal signals was significantly higher in the aggression context than the reference category (1.67 ± 0.34; $\chi^2_{12} = 16.90$, $P < 0.001$; Fig. 4). Effects of nonsignificant key predictors and control variables are provided in Appendix 3, Table A9.

Table 6
Individual mean number (SD) of multimodal combinations produced per communicative interaction across age categories and contexts

Age category (*N*)	MM combination	Context			
		Feeding	Play	Aggression	Other
Infant (8)	FV	0	4.30 (5.12)	0	0
	FG	0.10 (0.32)	2.50 (3.17)	0.10 (0.32)	0.50 (1.58)
	GV	0.60 (0.84)	0	0.10 (0.32)	0.10 (0.32)
	FGV	0.10 (0.32)	0	0.20 (0.42)	0
Juvenile (7)	FV	0.63 (1.21)	2.43 (3.95)	0 .83 (0.98)	0.00 (0.00)
	FG	0.83 (1.17)	6.71 (6.10)	0.33 (0.52)	0.29 (0.49)
	GV	1.86 (2.54)	0	1.00 (1.55)	0.86 (1.21)
	FGV	0.43 (0.79)	0	0.17 (0.41)	0.29 (0.76)
Early adolescent (11)	FV	0.36 (0.52)	1.64 (1.75)	1.50 (1.20)	0.36 (0.32)
	FG	0.09 (0.32)	0.45 (0.71)	0.73 (0.92)	0.82 (0.85)
	GV	0.8 (1.85)	0.18 (0.63)	2.55 (5.23)	0.36(0.32)
	FGV	0.18 (0.63)	0.18 (0.63)	0.73 (1.75)	0.09 (0.32)

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

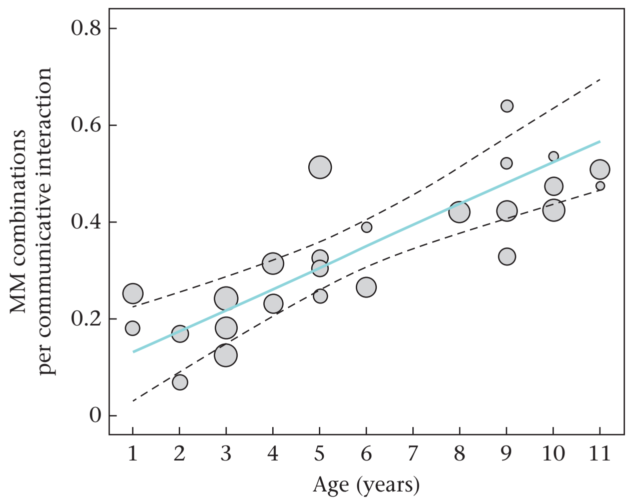

Figure 1. Effect of individual age (years) on the frequency of multimodal (MM) combination production in immature chimpanzees ($N = 28$). The mean number of MM combinations produced by each focal individual is shown. Area of the dots reflects the variation in sample size for each individual for each year of age. The solid line and dashed lines represent the fitted GLMM and 95% confidence intervals, respectively.

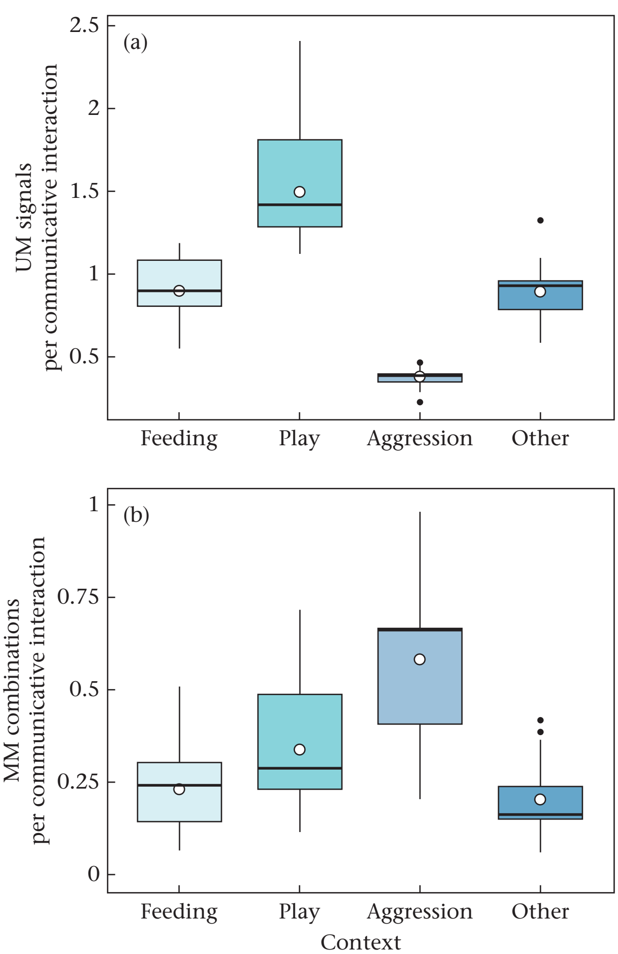

Figure 2. Effect of communicative context on the frequency of (a) unimodal (UM) signals and (b) multimodal (MM) combination production in immature chimpanzees ($N = 28$). Box plots show the mean number per context (open circles), median (horizontal lines), quartiles (boxes), percentiles (2.5 and 97.5%, vertical lines) and outliers (dots).

Variation in Frequency of Different UM 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 711 data points across all 28 individuals. Overall, the full models explained significantly more variation in the frequency of unimodal facial expression and gesture production than the null models (LRT comparing the full and null model for signal frequency: facial expressions: $\chi^2_7 = 360.28$, $P < 0.001$; gestures: $\chi^2_7 = 99.56$, $P < 0.001$). In neither model was the interaction term between age and context significant.

In terms of signal modality, age (year) did not have a significant effect on the frequency of unimodal facial (-0.10 ± 0.07; $\chi^2_{12} = 1.55$, $P = 0.21$) or gestural (-0.000 ± 0.06; $\chi^2_{12} = 0.000$, $P = 0.99$) 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 (2.43 ± 0.27; $\chi^2_{12} = 168.77$, $P < 0.001$; Fig. 5a) in our sample immature chimpanzees. In contrast unimodal gestures were produced significantly less frequently in the play context (-0.91 ± 0.13; $\chi^2_{12} = 48.97$, $P < 0.001$) and in the aggression context (-0.57 ± 0.23; $\chi^2_{12} = 6.72$, $P = 0.010$; Fig. 5b). We found a significant effect of the control factors of group and observation year in both models. Effects of all control variables and nonsignificant key predictors are provided in Appendix 3, Table A10.

Variation in Frequency of Different MM Combinations

Lastly, we intended to test whether variation in the frequency of different combinations of multimodal combinations were affected

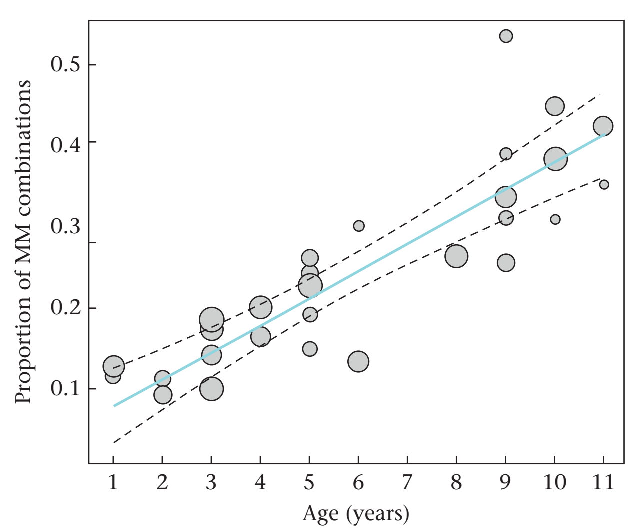

Figure 3. Effect of individual age (years) on the proportion of multimodal (MM) combinations produced by immature chimpanzees ($N = 28$). The proportional production of MM combinations in relation to unimodal signals is shown for each focal individual. Area of the dots reflects the variation in sample size for each individual for each year of age. The solid line and dashed lines represent the fitted GLMM and 95% confidence intervals, respectively.

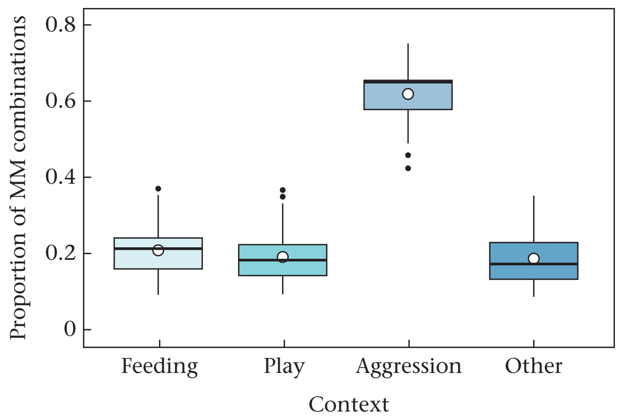

Figure 4. Effect of communicative context on the proportional production of multimodal (MM) combinations relative to unimodal signals in immature chimpanzees ($N = 28$). Box plots the mean proportion per context (open circles), median (horizontal lines), quartiles (boxes), percentiles (2.5 and 97.5%, vertical lines) and outliers (dots).

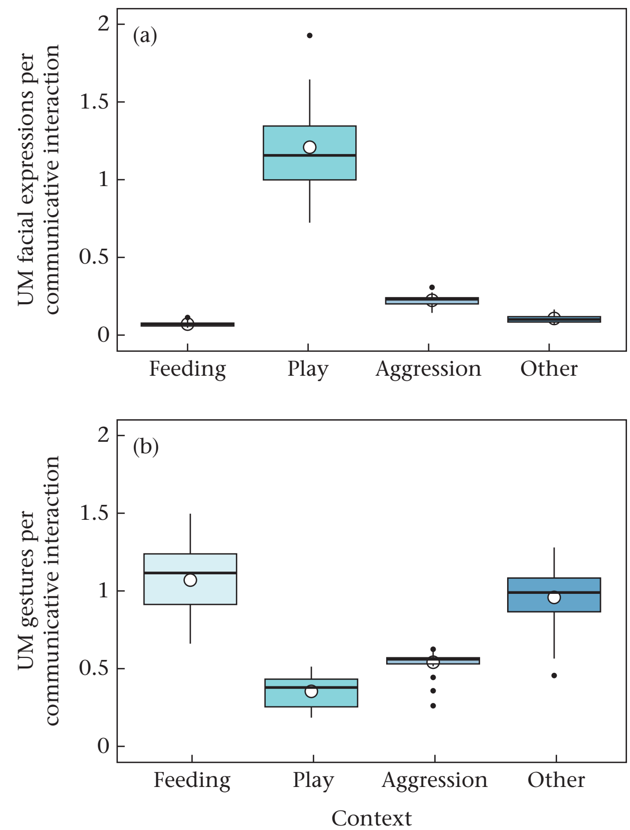

Figure 5. Effect of communicative context on the frequency of (a) facial and (b) gestural unimodal (UM) signal production in immature chimpanzees ($N = 28$). Box plots show the mean number per context (open circles), median (horizontal lines), quartiles (boxes), percentiles (2.5 and 97.5%, vertical lines) and outliers (dots).

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. Fig. 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 6.

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 (26/28 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. Gillespie-Lynch et al., 2014; Iverson, 2010), 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 recombinable signals. Indeed, while we did not analyse the distributions of different combination types here, descriptively we found that early adolescents in our sample

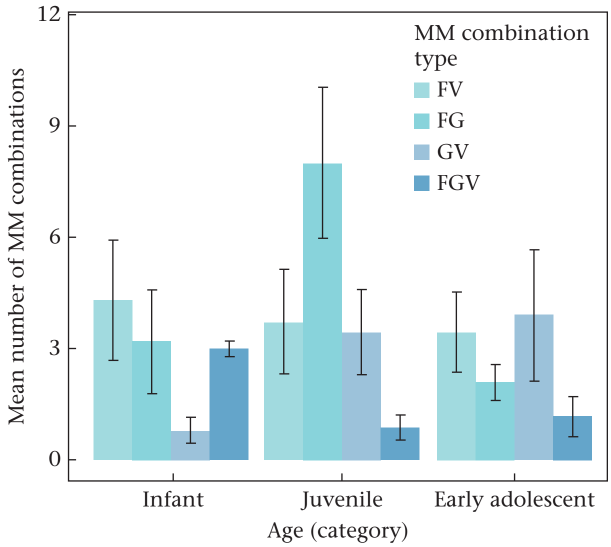

Figure 6. Variation in mean number of multimodal (MM) combinations produced across age categories of immature chimpanzees ($N = 28$). Error bars represent mean ± 1 SE. F = facial expression; G = gesture; V = vocalization.

produced the largest number of different multimodal combinations (see Table 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; 2015; 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 (Pusey, 1990; van Lawick-Goodall, 1968). 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 recombinable 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. 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, we found that facial expressions were the most frequent unimodal signal type produced in play, again due to the high number of playful open mouth faces. Unimodal gestures were observed significantly less, however, but in this context after facial–vocal combinations (i.e. open mouth face + laughter), facial–gesture combinations were most commonly observed (see Appendix 3, Table A7). 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 prefeeding 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; DeTroy 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.

Conclusion

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. Further, 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.

Author Contributions

E. Doherty: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Funding acquisition, Data collection, Visualization, Writing – original draft, Writing – review & editing. **Z. Clay**: Conceptualization, Investigation, Methodology, Supervision, Funding acquisition, Writing – original draft, Writing – review & editing. **M. Davila-Ross**: Conceptualization, Investigation, Methodology, Supervision, Funding acquisition, Writing – original draft, Writing – review & editing.

Data Availability

Data used in these models are available at https://data.mendeley.com/datasets/8wdrjdt376/2.

Declaration of Interest

The authors confirm no conflicts of interest.

Acknowledgments

We thank the staff members of the Chimfunshi Wildlife Orphanage and the Chimfunshi Research Advisory Board for their support. We also thank Hélène Chotard, Mackenzie Hewes, Sofia Vickers and Claire Barrault for contributing to the data collection. This research was supported by grants from the British Association of Biological Anthropology and Osteoarchaeology (BABAO), the Lucie Burgers Foundation for Comparative Behavioural Research (Arnhem, the Netherlands) and a research support training grant from Durham University (Durham Doctoral Scholarship). Research was approved by the Animal Welfare Ethical Review Board of Durham University. We are very grateful to the editor and two anonymous referees for their helpful comments on the manuscript.

References

Acredolo, L., & Goodwyn, S. (1988). Symbolic gesturing in normal infants. *Child Development, 59*(2), 450–466. https://doi.org/10.2307/1130324

Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour, 49*(3), 227–267. https://doi.org/10.1163/156853974x00534

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. https://doi.org/10.1002/ajp.22189

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. https://doi.org/10.1007/978-1-4614-0176-6_8

Bard, K. A., & Leavens, D. A. (2014). The importance of development for comparative primatology. *Annual Review of Anthropology, 43*, 183–200.

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.

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. https://doi.org/10.1002/ajpa.23325

Bretherton, I., & Bates, E. (1979). The emergence of intentional communication. *New Directions for Child and Adolescent Development, 1979*(4), 81–100. https://doi.org/10.1002/cd.23219790407

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. *R Journal, 9*(2), 378–400. https://doi.org/10.3929/ethz-b-000240890

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. https://doi.org/10.1007/s10071-017-1096-4

Call, J., & Tomasello, M. (2007). *The gestural communication of apes and monkeys*. Taylor & Francis/L. Erlbaum. https://doi.org/10.4324/9781003064541.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

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. https://doi.org/10.1007/s10071-014-0827-z

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), Article e23339. https://doi.org/10.1002/ajp.23339

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), Article 102175.

Esteve-Gibert, N., & Prieto, P. (2014). Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Communication, 57*, 301–316. https://doi.org/10.1016/j.specom.2013.06.006

Faraway, J. J. (2016). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. CRC Press.

Fitch, W. (2017). Empirical approaches to the study of language evolution. *Psychonomic Bulletin & Review, 24*(1), 3–33. https://doi.org/10.3758/s13423-017-1236-5

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. https://doi.org/10.1007/s00265-010-1038-5

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. https://doi.org/10.1016/j.anbehav.2016.12.018

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. https://doi.org/10.1111/brv.12535

Frohlich, M., & van Schaik, C. P. (2018). The function of primate multimodal communication. *Animal Cognition, 21*(5), 619–629. https://doi.org/10.1007/s10071-018-1197-8

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*(3), 483–500. https://doi.org/10.1007/s10071-015-0948-z

Frohlich, M., Wittig, R. M., & Pika, S. (2019). The ontogeny of intentional communication in chimpanzees in the wild. *Developmental Science, 22*(1), Article e12716. https://doi.org/10.1111/desc.12716

Genty, E. (2019). Vocal–gestural combinations in infant bonobos: New insights into signal functional specificity. *Animal Cognition, 22*(4), 505–518. https://doi.org/10.1007/s10071-019-01267-0

Genty, E., Clay, Z., Hobaiter, C., & Zuberbühler, K. (2014). Multi-modal use of a socially directed call in bonobos. *PLoS One, 9*(1), Article e84738. https://doi.org/10.1371/journal.pone.0084738

Genty, E., Neumann, C., & Zuberbühler, K. (2015). Bonobos modify communication signals according to recipient familiarity. *Scientific Reports, 5*(1), 1–10. https://doi.org/10.1038/srep16442
Ghazanfar, A. A. (2013). Multisensory vocal communication in primates and the evolution of rhythmic speech. *Behavioral Ecology and Sociobiology, 67*, 1441–1448. https://doi.org/10.1007/s00265-013-1491-z
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. https://doi.org/10.3389/fpsyg.2014.01228
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), Article e2004825. https://doi.org/10.1371/journal.pbio.2004825
Hobaiter, C., & Byrne, R. W. (2011). Serial gesturing by wild chimpanzees: Its nature and function for communication. *Animal Cognition, 14*(6), 827–838. https://doi.org/10.1007/s10071-011-0416-3
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. https://doi.org/10.1007/s00265-017-2325-1
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. https://doi.org/10.1017/S0305000909990432
Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science, 16*(5), 367–371. https://doi.org/10.1111/j.0956-7976.2005.01542.x
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.
Kojima, S. (2008). Early vocal development in a chimpanzee infant. In T. Matsuzawa (Ed.), *Primate origins of human cognition and behavior*. Springer. https://doi.org/10.1007/978-4-431-09423-4_9.
van Lawick-Goodall, J. (1968). The behaviour of free-living chimpanzees in the Gombe Stream Reserve. *Animal Behaviour Monographs, 1*, 161–311.
Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651), Article 20130302.
Liebal, K., Waller, B. M., Slocombe, K. E., & Burrows, A. M. (2014). *Primate communication: a multimodal approach*. Cambridge University Press.
Luef, E. M., & Pika, S. (2017). Reciprocal greeting in chimpanzees (*Pan troglodytes*) at the Ngogo community. *Journal of Neurolinguistics, 43*, 263–273. https://doi.org/10.1016/j.jneuroling.2016.11.002
Mundry, R. (2014). Statistical Issues and Assumptions of Phylogenetic Generalized Least Squares (PGLS). In L. Garamszegi (Ed.), *Modern phylogenetic comparative methods and their application in evolutionary biology* (pp. 131–153). Springer. https://doi.org/10.1007/978-3-662-43550-2_6.
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. https://doi.org/10.1044/2018_JSLHR-L-17-0402
Nishida, T., Kano, T., Goodall, J., McGrew, W. C., & Nakamura, M. (1999). Ethogram and ethnography of Mahale chimpanzees. *Anthropological Science, 107*(2), 141–188. https://doi.org/10.1537/ase.107.141
Oña, L. S., Sandler, W., & Liebal, K. (2019). A stepping stone to compositionality in chimpanzee communication. *PeerJ, 7*, Article e7623. https://doi.org/10.7717/peerj.7623
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. https://doi.org/10.1111/j.1439-0310.2004.01035.x
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. https://doi.org/10.1007/s10764-005-0724-z
Parr, L. A., Waller, B. M., Vick, S. J., & Bard, K. A. (2007). Classifying chimpanzee facial expressions using muscle action. *Emotion, 7*(1), 172. https://doi.org/10.1037/1528-3542.7.1.172
Partan, S. R. (2013). Ten unanswered questions in multimodal communication. *Behavioral Ecology and Sociobiology, 67*, 1523–1539.
Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science, 283*(5406), 1272–1273. https://doi.org/10.1126/science.283.5406.1272
Partan, S. R., & Marler, P. (2005). Issues in the classification of multimodal communication signals. *American Naturalist, 166*(2), 231–245. https://doi.org/10.1086/431246
Plooij, F. X. (1984). *The behavioral development of free-living chimpanzee babies and infants*. Monographs on infancy. Ablex.
Pollick, A. S., & De Waal, F. B. (2007). Ape gestures and language evolution. *Proceedings of the National Academy of Sciences, 104*(19), 8184–8189. https://doi.org/10.1073/pnas.0702624104
Pollick, A. S., Jeneson, A., & de Waal, F. B. (2008). Gestures and multimodal signaling in bonobos. In T. Furuichi, & J. Thompson (Eds.), *The bonobos: Behavior, ecology, and conservation* (pp. 75–94). Springer. https://doi.org/10.1007/978-0-387-74787-3_5.
Pusey, A. E. (1983). Mother-offspring relationships in chimpanzees after weaning. *Animal Behaviour, 31*(2), 363–377. https://doi.org/10.1016/S0003-3472(83)80055-4
Pusey, A. E. (1990). Behavioural changes at adolescence in chimpanzees. *Behaviour, 115*(3–4), 203–246. https://doi.org/10.1163/156853990X00581
Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for biologists*. Cambridge University Press.
Ron, T., & McGrew, W. (1988). Ecological assessment for a chimpanzee rehabilitation project in Northern Zambia. *Primate Conservation, 9*, 37–41.
Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution, 1*(2), 103–113. https://doi.org/10.1111/j.2041-210X.2010.00012.x
Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology, 20*(2), 416–420. https://doi.org/10.1093/beheco/arn145
Schneider, C., Call, J., & Liebal, K. (2012a). Onset and early use of gestural communication in nonhuman great apes. *American Journal of Primatology, 74*(2), 102–113. https://doi.org/10.1002/ajp.21011
Schneider, C., Call, J., & Liebal, K. (2012b). 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. https://doi.org/10.1007/s10764-011-9570-3
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. https://doi.org/10.1016/j.anbehav.2011.02.002
Slocombe, K., & Zuberbühler, K. (2010). Vocal communication in chimpanzees. In E. V. Lonsdorf, S. R. Ross, & T. Matsuzawa (Eds.), *The mind of the chimpanzee: Ecological and experimental perspectives* (pp. 192–207). University of Chicago Press. https://doi.org/10.7208/9780226492810-018.
Smith, W. J. (1977). *The behaviour of communicating: An ethological approach*. Harvard University Press.
Taglialatela, J. P., Russell, J. L., Pope, S. M., Morton, T., Bogart, S., Reamer, L. A., … Hopkins, W. D. (2015). Multimodal communication in chimpanzees. *American Journal of Primatology, 77*(11), 1143–1148.
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), Article e23249. https://doi.org/10.1002/ajp.23249
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. https://doi.org/10.1111/brv.12289
Townsend, S. W., Watson, S. K., & Slocombe, K. E. (2020). Flexibility in great ape vocal production. In L. M. Hopper, & S. R. Ross (Eds.), *Chimpanzees in context: A comparative perspective on chimpanzee behavior, cognition, conservation, and welfare* (pp. 260–280). University of Chicago Press. https://doi.org/10.7208/9780226728032-014.
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 Schaik, C. P. (2016). *The primate origins of human nature*. J. Wiley.
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. https://doi.org/10.1007/s10919-006-0017-z
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), Article 20130292.
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. https://doi.org/10.1016/j.jhevol.2017.10.010
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. https://doi.org/10.1111/j.1439-0310.2004.01045.x
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), Article 147470491301100305. https://doi.org/10.1177/147470491301100305
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. https://doi.org/10.1016/j.anbehav.2016.10.024

Appendix 1

Table A1
Full group size information in 2017 and 2021 field periods

	Group number		
Observation year	1	2	4
2017	25	52	12
2021	28	58	NA

NA: individuals from group 4 were not observed in 2021.

Table A2
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	188	50	65	306 (27, 10.97 ± 9.63)
Play	670	158	80	908 (27, 32.42 ± 31.89)
Aggression	49	73	67	189 (22, 6.32 ± 11.63)
Access[a]	17	9	5	31 (10, 1.04 ± 1.96)
Affiliation[a]	53	14	13	79 (22, 2.82 ± 2.73)
Grooming[a]	24	3	13	40 (11, 1.39 ± 3.29)
Rest[a]	26	6	6	38 (16, 1.25 ± 1.55)
Sexual[a]	25	2	3	30 (8, 1.07 ± 3.22)
Travel[a]	6	1	3	10 (5, 0.38 ± 0.81)

The 'other' signal type includes fixed multimodal, unimodal and multimodal sequences.
[a] Included in 'other' context category.

Appendix 2

Table A3
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	437	86	2	94	21
Juvenile	428	113	2	42	15
Early adolescent	195	117	9	46	21
Total	1058	316	13	182	55

This study focused on 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; Oña et al., 2019; Wilke et al., 2017; Frohlich 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 signals coded in each signal category is provided in Table A3.

Appendix 3

Table A4
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[a]	Lips are funnelled, parted and partially retracted. Mouth corners are pushed forward and the mouth is partially open
Scream face[a]	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[a]	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).
[a] Facial expressions produced in combination with associated vocal components were coded as fixed multimodal (facial–vocal) signals.

Table A5
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]	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[a]	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[a]	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).
[a] Vocalizations produced in combination with associated facial component were coded as fixed multimodal (facial–vocal) signals.

Table A6
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
Bite	Teeth are pressed into the skin of recipient's body
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 audible 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 audible
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
Hip thrust	Sitting, crouching or standing, the hips are thrust forward (single or repeated)
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 audible 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 audible '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 audible contact with a surface (2 feet together or alternately)
Stomp other	As Stomp but contact is with recipient's body
Tandem walk	Subject positions arm over the body of the recipient and both walk forward while maintaining position
Tap other	Single or multiple movements of the arm from the wrist or elbow with short but firm contact on a recipient's body
Throw object	Object is moved and released so that there is displacement through the air after release
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 A7
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.

Table A8
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

Table A9
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

Table A10
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

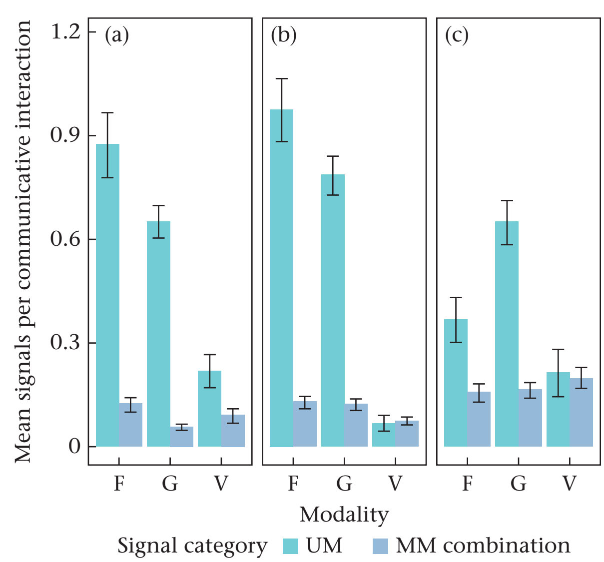

Figure A1. The mean frequency per communicative interaction of facial expressions (F), gestures (G) and vocalizations (V) produced as unimodal (UM) signals and as part of multimodal (MM) combinations in (a) infant, (b) juvenile and (c) early adolescent chimpanzees. Error bars represent mean ± 1 SE.