# Production of and responses to unimodal and multimodal signals in wild chimpanzees, *Pan troglodytes schweinfurthii*

# Claudia Wilkea, Eithne Kavanagha, Ed Donnellana,b, Bridget M. Wallerc, Zarin P. Machandad,e, Katie E. Slocombea

# a Department of Psychology, University of York, York, U.K.

# b Department of Psychology, University of Sheffield, Sheffield, U.K.

# c Department of Psychology, University of Portsmouth, Portsmouth, U.K.

# d Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, U.S.A.

# e Department of Anthropology, Tufts University, Medford, MA, U.S.A.

# Received 10 June 2016

# Initial acceptance 26 July 2016

# Final acceptance 12 September 2016

# MS number 16-00514

# Correspondence: K. Slocombe, Department of Psychology (office B201), University of York, Heslington, York YO10 5DD, U.K.

# E-mail address: [Katie.slocombe@york.ac.uk](mailto:Katie.slocombe@york.ac.uk)

# ABSTRACT

# Animals communicate using a vast array of different signals in different modalities. For chimpanzees, vocalizations, gestures and facial expressions are all important forms of communication, yet these signals have rarely been studied together holistically. The current study aimed to provide the first comprehensive repertoire of flexibly combined (‘free’) multimodal (MM) signals, and assess individual and contextual factors influencing production of, and responses to, unimodal (UM) and MM signals in wild chimpanzees. In total, 48 different free MM signals were produced. MM signals were produced at a significantly lower rate than UM signals, but 22 of 26 focal animals were observed to produce free MM signals. The relative production rates of different types of UM and MM signals differed significantly between the behavioural contexts investigated, showing flexible use of signals across contexts. In contrast, individual factors such as age, sex or rank of signaller did not appear to influence the type of signal produced or the likelihood of eliciting a response. Finally, we compared recipient responses to free MM grunt-gesture signals and matched UM component signals and found that these MM signals were more likely to elicit a response than a grunt alone, but were as likely to elicit a response as the gesture alone. The overall findings point to a widespread capacity for wild chimpanzees to flexibly combine signals from different modalities and highlight the importance of adopting a multimodal approach to studying communication.

# KEY WORDS

# chimpanzee, communication, multimodal, recipient response, signal production

# INTRODUCTION

Despite most animals producing multimodal (MM) signals (Hebets & Papaj, 2005; Rowe, 1999), researchers often focus on a single signal type (e.g. vocalizations), to the exclusion of all others. Reliance on such a unimodal (UM) approach to communication is particularly prevalent in nonhuman primate (primate) communication research; however, this approach unfortunately makes comparisons across modalities difficult and biases our understanding of the characteristics of signals in different modalities (Liebal, Waller, Slocombe & Burrows, 2013; Slocombe, Waller, & Liebal, 2011). Moreover, the MM signals that most animals produce are not captured by unimodal methods, and an important aspect of potential complexity in animal signalling may be lost as a consequence (Partan & Marler, 1999). Thus, we advocate that a MM approach that simultaneously investigates UM and MM signals using comparable methods is necessary to gain a comprehensive understanding of communication in any given species.

There are, however, some discrepancies and disagreements in the literature as to the definition of MM signals. In this paper we focus on ‘dynamic’ signals that ‘have a limited duration and require an action by the signaller to initiate (turn ‘on’) and to terminate the signal’, as this differentiates these signals from ‘state’ signals, which have static features that cannot be ‘turned off’, such as feather coloration (Smith & Evans, 2013, p. 1390). In terms of modality, while we acknowledge contrasting definitions in the literature (e.g. Higham & Hebets, 2013), we adopt the definition advocated by Waller, Liebal, Burrows and Slocombe, (2013). Rather than determining modality based on the sensory channels through which a signal is sent, such as auditory or visual signals, we use the term to refer to the type of communicative act commonly described in the literature in a given species (e.g. gestures, vocalizations and facial expressions in chimpanzees). Waller et al. (2013) argued that different cognitive processes or mechanisms may underlie different communicative acts, even if produced through the same sensory channel (such as gestures and facial expressions), and a single act can often produce sensory information through different channels (e.g. hand-clapping produces audio and visual output). Equally, it is important to distinguish between ‘fixed’ and ‘free’ MM signals. Fixed signals (Smith, 1977) are those whose component signals are necessarily combined due to the mechanics of signal production (e.g. a ‘pant hoot face’ necessarily accompanies a ‘pant hoot’ vocalization in chimpanzees). Conversely, free (also referred to as ‘flexible’ or ‘fluid’) MM signals are those whose components may be produced separately or combined flexibly with other signals (Tomasello, 2008). Finally, there is variation in the literature as to how MM signals are operationally identified. While fixed MM signals necessarily occur simultaneously, when considering free MM signals, some studies have looked for temporal overlap between signals (Partan, Larco & Owens, 2009), while others allow a margin of up to 10 s between the individual signals comprising a MM signal (Pollick & de Waal, 2007).

MM signal production has been reported in numerous taxa as diverse as ants (Uetz & Roberts, 2002), monkeys (Partan, 2002) and cowbirds (Cooper & Goller, 2004), and can involve the combination of a variety of different signals, such as seismic and visual signals (Hebets, 2008), or vocal and visual signals (de Luna, Hoedl & Amezquita, 2010; Partan, Larco & Owens, 2009). MM signals have been reported across a range of contexts, including alarm behaviour (e.g. Partan, Larco & Owens, 2009), aggressive interactions (e.g. Schwartz, 1974) and courtship (e.g. Hebets & Uetz, 1999). Several scientists have suggested that MM signalling can have several advantages over UM signalling for both producer and receiver, including increased signal detection and memorability, disambiguation of signals and allowing for more information to be transmitted (Liebal et al., 2013; Partan & Marler, 1999; 2005; Rowe, 1999).

In line with a framework offered by Partan and Marler (1999), the function of a MM signal can be determined by comparing recipient responses to the MM signal and the UM components in isolation. In the case of fixed vocal-visual MM signals, this has often been determined through careful experiments that used playbacks for vocal signals and animated models to test responses to visual signals. Although experiments remain the best way to study MM signal function and have been applied to free MM signals (Partan, Larco & Owens, 2009; 2010), the function of these signals can also be examined by collection of careful observational data on recipient responses to the MM signal and its component parts when produced unimodally. Broadly, MM signals can be categorized into (1) redundant combinations where recipients produce the same response to the component UM signals and the MM signal, but the response to the MM signal may be enhanced, and (2) nonredundant combinations where recipients produce different responses to the component UM signals, with possibilities for the responses to the MM signal to be different from those to the UM components (emergence) or more similar to those to one of the UM signals (dominance). To date, although MM signals are well documented in the animal kingdom, and have been rigorously investigated with elegant experiments in a number of nonprimate species, there is a lack of comparable investigation into MM communication in primate species (Liebal et al. 2013).

Understanding the communicative abilities of primates is not only important for establishing a window into their complex social world and cognitive abilities, but also for understanding human language evolution. Mapping out the differences and similarities in communicative abilities of humans and our closest living relatives may help us discern which are the derived, uniquely human aspects of language and which may have built on abilities already present in common ancestors with extant primates. In addition, characteristics of primate vocal and gestural communication provide key lines of evidence for theories concerning whether language has vocal or gestural origins (Slocombe, Waller & Liebal, 2011). Among the primates, chimpanzees, our closest living relatives, provide the best model of what our last common ancestor might have been capable of, and thus play a critical role in informing debates on the evolutionary origins of human language (Hayashi, 2007; Watson et al, 2015; Schel, Machanda, Townsend, Zuberbühler & Slocombe, 2013; Taglialatela, Russell, Schaeffer & Hopkins, 2011; Hobaiter & Byrne, 2011a).

For chimpanzees, vocalizations, gestures and facial expressions are all important forms of communication, and previous UM research on these different types of signals have investigated characteristics such as intentionality (e.g. Leavens, Hopkins & Thomas, 2004; Schel, Townsend, Machanda, Zuberbühler, & Slocombe, 2013; Hopkins, Taglialatela & Leavens, 2011), referentiality (e.g. Slocombe & Zuberbühler, 2005; 2006; Crockford, Wittig & Zuberbühler, 2015), flexible use across contexts (e.g. Hobaiter & Byrne, 2011) and audience effects (e.g. Gruber & Zuberbühler, 2013; Leavens, Hopkins & Bard, 1996; Kalan & Boesch, 2015; Schel, Machanda, et al, 2013; Slocombe & Zuberbühler, 2007; Slocombe et al. 2010; Townsend & Zuberbühler, 2009). On the surface, this UM work indicates that gestures, vocalizations and facial expressions differ in terms of these characteristics; however, few studies have attempted to examine these characteristics in a comparable manner in multiple modalities, so such conclusions may be premature (Slocombe et al., 2011). One study that has successfully examined different types of signal within a single experimental paradigm explored whether captive chimpanzees could selectively produce a signal appropriate to the attentional state of a human. Leavens, Russell and Hopkins (2010) showed that chimpanzees, while begging from a human experimenter, used more visual gestural signals when the researcher was facing towards them, and more tactile and vocal signals when they were facing away.

Despite the wealth of research on the production of vocal, gestural and facial signals in isolation, the combination of these signal types into MM signals in chimpanzees is virtually unexplored (Liebal et al., 2013; Slocombe, Waller & Liebal, 2011). Important exceptions to this include an experimental study probing recipient integration of signals from different modalities, which revealed that chimpanzees can cross-modally match facial expressions and vocalizations (Parr, 2004). In addition, Parr found that either the vocal or facial components were more salient to the chimpanzees depending on the signal type (e.g. the vocal component of a pant hoot signal was more salient than the facial component). From a production perspective, a recent study by Taglialatela et al. (2015) indicated that approximately 50% of captive chimpanzee vocalizations were accompanied by nonvocal signals (e.g. gestures, fear grimace) or behaviours (e.g. chase, play), and that these combined signals were more likely to be directed towards another individual than vocal signals alone. This indicates that chimpanzees may use signal combinations from different modalities strategically to meet specific sociocommunicative goals. Focusing on the combination of gestural signals with vocal or facial signals in captive chimpanzees, Pollick and de Waal (2007) found 21% of chimpanzee signals were MM. However, the operational definition of MM signals probably captured MM sequences as well as signals, as signals occurring within 10 s of each other were counted as MM signals. Perhaps surprisingly, MM signals were not found to be more effective in eliciting a response than UM signals. However, unfortunately, this study’s findings are difficult to interpret as the analyses also suffer from pseudoreplication (Waller et al. 2013). Despite variation in how these two studies define a MM signal, it seems that in captivity, where visibility of group members is usually excellent, vocal, gestural and facial signals may be commonly combined into MM sequences or signals. The degree to which chimpanzees produce MM signals in their visually dense natural habitat, and whether in a wild setting MM signals are more effective at eliciting responses than UM signals, remains unknown. In addition, despite free MM signals having the potential to generate new meaning (emergent function; Partan & Marler, 1999) and to indicate cognitive complexity relevant to a language evolution perspective (Slocombe, Waller & Liebal, 2011), we are currently lacking a MM repertoire and an understanding of how common and varied such free combinations may be.

In this study we attempted to address these issues and systematically investigated the UM and MM communication of wild chimpanzees, by taking an integrated MM approach. We considered MM signals temporally overlapping combinations of vocal, gestural and facial signals. We aimed to provide the first MM signal repertoire, understand the individual and contextual factors that affect UM and MM signal production, and compare the recipient responses to MM and matched UM signals.

In terms of signal production, we predicted that the rate of UM signal production would be significantly higher than that of MM signals. Furthermore, we expected MM rates may be lower than those found in captivity, due to adoption of stricter criteria and the more restricted transmission of visual signals in a dense forest environment. Second, in terms of individual factors, we expected that younger, female or more subordinate individuals may show higher rates of MM than UM signals, as they may need to show more elaboration in signalling in order to elicit responses than older, male, dominant individuals. Third, focusing on UM signals, given that captive chimpanzees modulate signal type depending on the recipient’s visual attention (Leavens, Russell & Hopkins, 2010), we predicted that relative rates of vocal, gestural and facial signals would vary with context, with higher rates of gestures and facial expressions in contexts where visual signals would be most visible for receivers (e.g. rest, groom).

In terms of recipient responses, we predicted that MM signals would elicit significantly higher proportions of recipient responses than matched UM signals, as increased recipient responses to MM signals have been repeatedly found in rigorous nonprimate studies (reviewed in Liebal et al., 2013), because recipients are more likely to detect and attend to these more elaborate and salient signals. We also predicted that recipient responses would be more likely when the signaller was more dominant and there were more recipients in the vicinity.

# METHODS

*Study Site and Subjects*

This study was carried out in Kibale National Park, located in western Uganda (0”13’ –0”41’N and 30”19’ –30”32’E) in 2013 –2015. A detailed description of the characteristics of the forest can be found in Chapman and Wrangham (1993). The study animals were a wild group of chimpanzees, the Kanyawara community. In 2013, the group comprised approximately 57 individuals (Muller & Wrangham, 2014), and occupied a home range of around 16.4 km² (Wilson, Kahlenberg, Wells & Wrangham, 2012). The community is entirely habituated and have been followed and studied regularly since 1987 by the Kibale Chimpanzee Project (Wrangham, Clark & Isabirye-Basuta, 1992; Georgiev et al., 2014). Specifically, the individuals included in this study were 13 males and 13 females, from 8 to 47 years old (see Table 1). These individuals were chosen on the basis that they were easy to find and follow, ensuring that as much high-quality focal time as possible could be collected for each individual. Dominance ranks were established by calculating a modified David’s score, MDS (de Vries, Stevens & Vervaecke, 2006), for all individuals for which long-term field assistant data on decided aggressive interactions and submissive pant grunt vocalizations were available (these data were unavailable for some younger individuals; their rank was noted as NA). MDS was calculated for males and females separately and all males were ranked above all females, as all of these males had dominated the females.

Table 1. ID, sex, age and rank of the 13 male and 13 female focal individuals

|  |  |  |  |
| --- | --- | --- | --- |
| ID | Sex | Age (years) | Rank |
| AJ | M | 39 | 4 |
| AL | F | 31 | 12 |
| AT | M | 14 | 7 |
| AZ | M | 9 | NA |
| BB | M | 47 | 5 |
| BO | M | 10 | NA |
| ES | M | 19 | 1 |
| LK | M | 31 | 3 |
| LN | F | 16 | 16 |
| ML | F | 16 | 14 |
| NP | F | 13 | 18 |
| OG | M | 12 | NA |
| OM | F | 8 | NA |
| OT | F | 15 | 19 |
| OU | F | 34 | 9 |
| PB | M | 18 | 6 |
| PO | F | 14 | 15 |
| TG | F | 33 | 10 |
| TJ | M | 18 | 2 |
| TS | F | 8 | NA |
| TT | M | 13 | NA |
| UM | F | 32 | 13 |
| UN | M | 9 | NA |
| WA | F | 22 | 17 |
| WL | F | 21 | 11 |
| YB | M | 40 | 8 |

Age in 2013, the first year of data collection. Rank order is based on the modified David’s score. NA indicates young individuals for whom these data were not available.

*Equipment*

All focal observational data were collected with a Panasonic HDC-SD90 camcorder, with a Sennheiser MKE 400 microphone attached. Recipient responses were recorded with a second camcorder: a Panasonic HDC-SD40. Videos were coded using Noldus Observer XT 10 event logging software (http://www.noldus.com/animal-behavior-research) for observational data.

*Ethical Note*

This study complied with the ASAB/ABS guidelines for the use of animals in research; ethical approval was granted by the Biology Ethics Committee (University of York). The Ugandan Wildlife Authority and the Ugandan National Council for Science and Technology granted permission to carry out the study in Uganda.

*Data Collection*

All data were collected February–May 2013 and June 2014 –March 2015, between 0800 and 1830 hours. Focal animal sampling (Altmann, 1974) was employed in order to collect observational data on the 26 focal individuals. Focal animals were only sampled once a day and were chosen in a way that maximized the quality and spread of data across target individuals. Initially, once a party of chimpanzees were located, the target animal with the best visibility for clear filming was chosen as the focal individual, but later on in the study period target individuals with the least focal time were prioritized.

**Focal animal signal production**

Focal samples consisted of 15 min of continuous video observation of one focal animal. The aim was to capture on video a complete view of all facial, gestural and vocal signals produced by the focal individual. Thus the camera was zoomed in as close as possible, while still capturing the whole body of the chimpanzee. The researcher commentated all vocalizations in real time, to ensure that even quiet vocalizations that could not be picked up by the microphone were recorded. If individuals moved out of sight and earshot during a focal observation, this time was coded as ‘out of sight’ (OOS) and excluded from any further analysis. Samples containing more than 10 min of OOS time were excluded from further analysis, meaning the analysed samples range from 5 to 15 min and all had a good level of visibility of the focal animal. As we were interested in social communication, only focal samples during which the focal individual was in a party (i.e. there were other independent individuals within 30 m; Slocombe et al, 2010) were included in this analysis. Thus this excluded times when mothers were alone, with only their infants, as mother – infant communication could not be examined in the majority of the focal individuals, who were not mothers.

**Recipient responses**

To collect the response of other individuals to any signals produced by the focal individual, a second researcher used a camcorder to capture the signals and behaviour of as many of the individuals closest to the focal animal (within 5 m) as possible.

*Video Coding*

Video coding with Observer XT 10 software was used to extract continuous details about the behavioural context and modality availability of the focal individual, as well as all their UM and MM signal production (see detailed definitions below). By coding the context and modality availability continuously we were able to calculate accurate rates of signal production in each context, as a function of the time each specific modality could be reliably detected. The types of behaviours elicited from individuals within 5 m of the focal individual, in response to each focal signal, were also coded from the videos (see detailed definitions of these responses below). Recipient responses were only coded for the 32 h of video data for which a second observer was present to capture these on a second video camera.

*Definitions*

**Behavioural contexts**

Eight behavioural contexts were defined and coded, but only four occurred frequently enough across focal animals to be examined further in terms of signal production rates (rest, groom, feed, travel; see Table 2). For these four contexts, the requisite behaviour had to continue for at least 20 s (a break of up to 5 s is permissible during this time), after which breaks of up to 15 s were permissible, as long as the individual always returned to the original behaviour. For the repertoire and number of instances of different MM combinations (see Table 5 below and Table S3 in the Supplementary material), all contexts were included i.e. all available video time, in order to establish a more comprehensive picture of the types of signals that the focal individuals were motivated to combine.

Table 2. Description of the behavioural contexts of the focal individual

|  |  |
| --- | --- |
| Behavioural context | Description |
| Rest | When the focal animal is sitting or lying down relatively still with eyes open, and for most of the time not feeding, grooming or playing. Also includes time spent self-grooming (attending to their own body/fur: combing through the fur or picking at the skin to remove dirt or parasites) |
| Feed | When the focal animal is collecting and eating, or extracting moisture from, food (e.g. leaves, bark, fruit, honey). It may move short distances in the process of doing this. If it resumes feeding after a period of chewing, this continues to be counted as feeding. If it chews for more than 3 min without resuming collecting and eating more food after this, this is coded as resting after 3 min of chewing |
| Groom with other | When the focal animal is attending to the body/fur of another individual: combing through the fur or picking at the skin to remove dirt or parasites |
| Travel | When the focal animal is walking or running for most of the time (may sometimes halt for brief periods). Movement associated with play (such as chasing), feeding (such as moving short distances for foraging), displaying or aggression is not regarded as ‘travel’ |
| Other | Includes sleep (eyes closed and no movement), social play (Nishida et al., 1999), aggression (includes threats, chasing, physical violence, etc.) and display (includes charging, body swaying, branch shaking, dragging and throwing, etc., see Nishida et al., 1999) |

Some behaviours could occur simultaneously; thus there was a hierarchy when coding, with the more active (generally also rarer) behaviour being given precedence: travel>feed, display>travel, play>travel, aggression>travel, aggression>display, play>feed, groom with other>self-groom. It was occasionally the case that an individual would be involved in an agonistic interaction and then rest, feed, travel or groom immediately afterwards. In these cases, behavioural contexts were still based on the current behaviour of the individual; thus rest, feed, groom and travel contexts also include postconflict periods.

**Modalities available**

This behavioural coding group was used to capture which type of signals produced by the focal individual could be coded reliably from the video at all times (see Supplementary Table S1). It was frequently the case that only signals in one or two modalities could be accurately captured due to the orientation of the focal animal (e.g. face may not be visible) or distance of observer to the chimpanzee (e.g. quiet vocalizations may not be detected). This was coded so it could be taken into account when calculating rates of signal production.

To be coded as a ‘modality available’, the requisite modality had to be available for at least 20 s (a break of up to 5 s is permissible during this time), after which breaks of up to 15 s were permissible, as long as the original modality then became available again. The exception to this rule was when the modality could not be seen for most of the time, but in the short period for which it was available, a signal was produced (for instance an individual’s face cannot be seen, they then turn around for 2 s, showing a ‘play face’, then turn away again); in this case it was coded as available for this short period, and the signal produced was also coded.

**Focal animal signals produced**

We coded all vocal, gestural (manual and nonmanual) and facial signals the focal individual produced (see Table 3). The duration of facial and gestural signals was coded; for vocalizations, which were commonly produced in bouts, the duration of the calling bout was recorded.Two or more vocalizations of the same type were coded as one continuous bout if they were produced within 10 s of one another (from the end of one to the beginning of the next). Eight different facial expressions were coded; these were based on the prototypical expressions described in Parr, Waller, Vick and Bard (2007), which are specific combinations of facial muscle movements (Action Units: ChimpFACS, Vick et al. 2007). The person coding the signals discussed exemplars with B.W. (certified FACS coder) prior to video coding in order to avoid any expressions that did not fit the prototypical descriptions. Forty common gestures were coded based on the repertoire proposed by Hobaiter & Byrne (2011a). Rare gestures were coded as ‘other manual gesture’ or ‘other nonmanual gesture’ and described in the notes section. Similarly, owing to the size limits of the coding scheme, some gestures were combined under an umbrella term, for instance ‘manual contact with another individual’ included touch, hand on, punch, push, slap, tap, poke, hit. Fourteen different vocalizations were coded based on the repertoire proposed by Slocombe and Zuberbühler (2010).

Table 3. The number and type of signals coded in each modality

|  |  |
| --- | --- |
| Signals coded in each modality | Signal types |
| Facial expressions (*N* = 8) | bared teeth display; play face; pant hoot face; scream face; alert face; pout; whimper face; ambiguous face |
| Vocalizations (*N* = 14) | Pant hoot; whimper; scream; squeak; bark; waa bark; cough; grunt; rough grunt; pant grunt; pant; alarm huu; laughter; soft hoo |
| Manual gestures (*N* = 20) | Brief manual contact with object or ground; manual contact with another individual; manually displace object; arm raise; arm shake; arm swing; arm wave; beckon; big loud scratch; clap; drum belly; embrace; hand fling; hand shake; hide face; leaf clip; mouth stroke; reach; shake hands; water splash |
| Nonmanual gestures (*N* = 20) | Bite; bow; dangle; feet shake; foot present; gallop; head nod; jump; kick; leg swing; look; object in mouth approach; present climb on me; present grooming; present sexual; roll over; rump rub; stomp; stomp other; walk stiff |

See Supplementary Table S2 for detailed descriptions of each signal type.

**Recipient response time and types**

Recipient responses were coded from the beginning of the focal individual signal until 20 s after the signal had finished, from individuals within 5 m of the focal individual. During the recipient response time the number and identity of the individuals within 5 m of the focal individual were recorded. If another signal occurred within the 20 s after the first signal then the recipient response time was cut short for the first, with this only lasting until the beginning of the next signal. Similarly, if the recipient response continued after the 20 s (for example the signal elicited a long bout of grooming), this was also only coded for up to 20 s after the end of the focal signal.

Recipient responses comprised four groups: signal responses (facial, vocal, gestural and MM), movements, negative and positive responses (see Table 4). Signals by other individuals were only coded as responses if the recipient’s facial expression or gesture was directed at the focal individual (as far as this was relevant and possible to discern). It was difficult to determine specifically to whom vocalizations were directed, so all vocalizations from recipients were counted as potential signals in response to the focal individual. Any signals or behaviours that were clearly in response to an unrelated signal or event were not coded as responses. For example, if the focal individual gave a big loud scratch (BLS) gesture, and immediately afterwards individuals in another party uttered pant hoots and an individual within 5 m of the focal animal replied with a pant hoot, the pant hoot was not coded as a response to the BLS. Similarly, only an active change in behaviourof the recipient was coded as a response. For instance, if another individual was already vocalizing, and then the focal individual produced a signal, and the other individual continued vocalizing as before, this was not counted as a response to the focal animal’s signal. Equally, ‘terminating’ behaviours were not coded, for example the cessation of playing or grooming. Behavioural responses (positive, negative) had to be directed towards the focal animal rather than a third party to be counted as a response to the focal animal’s signal.

Table 4. Description of the types of recipient responses coded

|  |  |
| --- | --- |
| Responses of recipients | Description |
| Facial, vocal, gestural or MM response | The facial expressions, vocalizations and gestures given by recipients were coded in the same way as those of the focal individual (see Table 3 and Supplementary Table S2) |
| Movement response | Movement responses were coded when a recipient was clearly moving directly towards, or directly away from, the focal individual, by at least 2 m (excluding occasions where the recipient was merely passing). This was coded at the point the movement began |
| Positive response | Positive responses from recipients included recipients grooming or playing (see description of these behaviours in Table 2) with the focal animal, or clearly desired responses e.g. ‘present climb on me’ is followed by the recipient climbing on the signaller. Grooming was coded as a response either when grooming was initiated, or the recipient changed where they were grooming (as was often the case when the focal individual presented a new body part) |
| Negative response | Negative responses from recipients included fearful or submissive responses (running away, a cowering body posture, showing fearful facial expressions, screaming), as well as mild (threats, nondirected display) or severe aggression (chasing, directed display, physical violence, see Nishida et al, 1999) |

*Calculation of signal production rates*

A total of 121 h of videos were coded. Of these, 111 h were in groom, rest, feed and travel contexts and thus were used for the calculation of UM and MM signal production rates. To ensure that signal production rates were representative of an individual’s behaviour, we set minimum amounts of time that an animal could have been observed to produce the relevant signal in key contexts in order to enter analyses.

For UM signals, rates were only calculated for UM facial, vocal and gestural signals for a specific context for an individual if they had at least 30 min of this modality available in rest, feed and groom contexts, and at least 20 min in the travel context. For example, to have a rate for facial expressions in the rest context, that individual must have at least 30 min of facial expressions available during rest. Mean UM rate was the average of facial, vocal, manual gesture and nonmanual gesture rates. Individuals must have contributed to all of these to have a mean UM rate calculated in a specific context. Those who did not meet this criterion had a missing value for UM rate in this context.

Rates were only calculated for MM combinations for a specific context for an individual if they had at least 15 min of this modality combination available in rest, feed and groom contexts, and at least 10 min in the travel context. For example, to contribute a rate for vocal-gestural signals in the travel context, that individual must have had at least 10 min of time where both vocalizations and gestures were simultaneously available in this context. Mean MM rate was the average of facial-gestural, vocal-gestural, fixed facial-vocal, free facial-vocal and facial-vocal-gestural rates. Individuals must have contributed at least three of these MM combination rates to have a mean MM rate calculated in a specific context. Those who did not meet this had a missing value for MM rate in this context.

Mean signal production rates for the group (as reported in descriptive statistics and figures in the Results) were calculated as a mean of all the individual mean production rates that contributed to a particular analysis.

To assess whether the number of individuals in the party affected signal production, we calculated the average number of individuals in the party present during the periods from which signal rates were calculated for each type of signal produced by each individual. First, the number of individuals in the party was recorded at the beginning and end of every video and these were averaged. Second, for each signal type for which a rate was calculated for an individual, we took the corresponding videos that had contributed to the calculation of that rate and calculated a mean from the average number of individuals in the party across those videos.

*Comparison of responses to MM signals and UM components*

# Most previous nonprimate research carried out within a MM framework has focused on fixed MM signals and/or signals produced only in one specific context, for example alarm behaviour (e.g. Partan et al., 2009) or courtship behaviour (e.g. Uetz, Roberts & Taylor, 2009). In contrast, the signals we investigated were free MM signals, which were produced across a range of contexts (see Supplementary Table S4). As context was shown to heavily influence signal production (see signal production results below), we endeavoured to control for this by matching UM and MM signals based on signaller identity and behavioural context of production. We consider such matching of MM signals and UM component signals to be critically important in order to understand the function of the signals.

We focused on the free MM signal produced most frequently by the largest number of individuals, where matched UM component signals were also frequently produced by the same individuals: the grunt + gesture signal (vocal-gestural combination).It was not possible to examine more MM signal combinations as no other type of free MM signal, with sufficient matched UM components, was produced by a sufficient number of individuals.

For each of the MM signals we identified component UM signals that were matched to the MM signal in terms of the behavioural context during production. Up to five UM grunt signals and five UM gesture signals were matched to each MM signal. Where possible we also matched the number of individuals present within 5 m of the focal individual. For instance, if the individual PO produced a grunt + present groom MM signal in a groom context, with two individuals within 5 m, the responses to this signal could be compared to the responses to a UM grunt vocalization from PO, in a groom context, with two individuals within 5 m of her, and a UM present groom gesture, in a groom context, with three individuals within 5 m of her.

*Intercoder Reliability*

To assess the intercoder reliability of the video coding, a second independent researcher also coded 6.5% of the videos (7.75 h, *N* = 31 videos each lasting 15 min from a total of 15 individuals), having been provided with comprehensive instructions. Cohen’s kappa was calculated; the mean Kappa value obtained was 0.81, indicating excellent levels of coder agreement (Fleiss, 1981). All reliability analyses were run using the Reliability Analysis function in Observer XT 10, which enables the comparison of two different Event Logs for one video.

*Data Analysis*

We constructed linear mixed models (LMMs) and generalized linear mixed-effects models (GLMMs) in order to test our hypotheses regarding signal production and recipient responses respectively. LMMs were used to investigate the influence of continuous and categorical variables on signal production rates, while GLMMs with a binomial error structure were used to investigate the influence of continuous and categorical variables on the occurrence of recipient responses (binary response variable: received one or more responses or no responses). Furthermore, because we had repeated sampling from the same individual, to control for pseudoreplication we fitted ‘individual’ as a random factor (Crawley, 2002) by conducting random intercepts models using the package lme4 (Bates & Maechler, 2009; https://cran.r-project.org/web/packages/lme4/index.html). We first assessed whether the full model could explain a significant amount of variation in the dependent variable, by comparing the full model to a null model containing just the intercept and random factors. To assess the significance of each explanatory variable or interaction term, we compared the full model with a reduced model excluding the variable or interaction of interest using a likelihood ratio test (Faraway, 2006). All models were run in R v. 2.15 (The R Foundation for Statistical Computing, Vienna, Austria, <http://www.r-project.org>).

As some data were not available for all individuals (e.g. dominance rank) or were missing in the majority of individuals due to methodology (e.g. facial expression was not available during travel, as the observer followed and filmed travelling chimpanzees from behind), we sometimes constructed several models in order to test our hypotheses thoroughly, and to maintain a high number of individuals in each model.

# RESULTS

# *MM Signals: Repertoire, Rates and Responses*

Overall, the results show that across rest, feed and groom contexts MM signals were rare relative to UM signals (see Fig. 1). Free MM signals were, however, produced by 22 of the 26 focal individuals, and we recorded a total of 48 different free MM signals, consisting of combinations that in total included six different facial expressions, nine different vocalizations and 16 different gestures (see detailed MM repertoire in Supplementary Table S3). Vocal-gestural combinations were the most common free MM signals recorded, and free facial-vocal the least (see Table 5). The frequency of different types of responses the various different categories of MM signal elicited from those within 5 m are also shown in Table 5. Vocal-gestural signals were the most likely to elicit any kind of response, and the most likely of all the signal combinations to elicit a positive response. In contrast fixed facial-vocal signals received the highest percentage of negative responses.

Figure 1. The mean signal production rate (per h) of UM and MM signals in the contexts rest, feed and groom. Error bars represent bootstrapped 95% confidence intervals, based on 1000 iterations. Error bars that do not overlap represent a significant difference. Mean rate and confidence intervals derived from data of *N* = 23 (UM rest, MM groom), *N* = 25 (MM rest), *N* = 11 (UM feed), *N* = 26 (MM feed), *N* = 9 (UM groom).

Table 5. Occurrences of different MM combinations and responses to these

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Across all eight contexts | |  | Across rest, feed, groom and travel contexts | | | | | |
|  | MM signal combination | No, of individuals | Total no. of occurrences (range) |  | No. of occurrences | F, V, G or MM signal response  (%) | Movement response  (%) | Positive response  (%) | Negative response  (%) | No response  (%) |
| Free | F-G | 14 | 47 (1-9) |  | 5 | 20 | 20 | 0 | 0 | 60 |
|  | F-V | 9 | 14 (1-4) |  | 8 | 25 | 0 | 0 | 0 | 75 |
|  | V-G | 15 | 53 (1-15) |  | 46 | 20 | 9 | 41 | 4 | 26 |
|  | F-V-G | 13 | 36 (1-12) |  | 12 | 8 | 17 | 0 | 0 | 75 |
| Fixed | F-V | 20 | 95 (1-11) |  | 57 | 35 | 4 | 5 | 12 | 56 |

The table shows the number of instances and number of individuals observed to produce different MM combinations, both free and fixed, across all eight behavioural contexts, with the range of number of occurrences a combination was produced by a single individual in parentheses. It also shows the number of occurrences where the signal was produced in rest, feed, groom or travel contexts with at least one recipient within 5 m, and of these, the percentage of these occurrences that elicited each of the four recipient response types, or no response. One signal could elicit several responses. Responses were recorded from the start of the signal until 20 s after the end of the signal. The table includes ambiguous signals, where the modality combinations were clear (e.g. facial-vocal signal) but at least one of the specific signal types could not be easily categorized given the signal repertoires used (Table 3). F: facial; V: vocal; G: gestural.

*Variation in MM signal production rates: free versus fixed*

As there is a key cognitive distinction between free MM signal combinations, where signals may be flexibly ‘mixed and matched’, and fixed MM signal combinations, which are necessarily combined, we investigated the individual and contextual factors that might influence the relative rates of these signals. We constructed a model to test whether variation in the mean MM signal production rate (signals/h) could be explained by interactions between the following fixed factors: type of MM signalling (fixed, free) and (1) context of production, (2) the mean number of individuals in the party, (3) the age of the signaller and (4) the sex of the signaller. The travel context was not included as most MM combinations included facial expressions, which could virtually never be captured during travel.

The full model comprised these interaction terms and the associated fixed factors. Individual identity was included as a random factor. The dependent variable was mean rate of MM signal production/h. We included 156 data points from 26 individuals in the model. Overall, the full model (*N* = 26 individuals) did not explain a significant amount of variation in MM signal production rates, compared to a null model (X211 = 17.06, *P* = 0.106), indicating that these factors and interactions did not account for significant variation in the MM signal production rates.

As rank was only available for older individuals, we ran a separate model to assess the effects of rank, by adding rank as a fixed effect and the interaction between rank and signal type to the full model specified above (*N* = 114 data points from 19 individuals). This version of the model that included rank (*N* = 19) confirmed that the overall model did not explain a significant amount of variation in MM signal production rates (X213 = 6.70, *P* = 0.917).

Rates of the different types of free MM signal combinations were too low and lacked sufficient variability (e.g. the majority of individuals had rates of 0 signals/h) to be subject to inferential statistics; however, Fig. 2 shows that there was interesting variation in the type of MM signals produced in rest, feed and groom contexts.

Figure 2. The MM signal production rate (per h) of facial-gestural (FG), vocal-gestural (VG), facial-vocal fixed (FV fixed), facial-vocal free (FV free) and facial-vocal-gestural (FVG) combinations in the contexts rest, feed and groom. Error bars represent bootstrapped 95% confidence intervals, based on 1000 iterations. Error bars that do not overlap represent a significant difference. MM rate and confidence intervals derived from data of *N* = 25 (facial-gestural, facial-vocal fix, facial-vocal free and facial-vocal-gestural rest), *N* = 26 (vocal-gestural rest, vocal-gestural, facial-vocal fix, facial-vocal-gestural feed and vocal-gestural groom), *N* = 23 (facial-vocal fix and facial-vocal flex groom). Missing bars occur when the MM rate was zero.

*Factors Affecting UM and MM Signal Production*

We constructed a model to test whether variation in the mean signal production rate (signals/h) could be explained by interactions between the following fixed factors: type of signalling (UM, MM) and (1) context of production (rest, feed, groom), (2) the mean number of individuals in the party, (3) the age of the signaller and (4) the sex of the signaller. The full model comprised these interaction terms and the associated fixed factors. Individual identity was included as a random factor. The dependent variable was rate of signal production/h. We included 117 data points from 26 individuals in the model. Note that for this model the travel context was excluded as no individuals had sufficient time for UM facial expressions or any MM combination involving facial expressions (i.e. facial-gestural, facial-vocal, facial-vocal-gestural) available in this context.

Overall, the full model (*N* = 26 individuals) explained a significant amount of variation in signal production rates, compared to a null model (X211 = 147.06, *P* < 0.001). Likelihood ratio tests revealed that there was a significant interaction between type of signal and context (X22 = 11.12, *P* = 0.004; Fig. 1). Figure 1 illustrates that signal production rates were significantly higher for UM signals than for MM signals in each context, but the difference between UM and MM rates was greatest in groom and rest contexts, compared to the feed context. No significant interactions between signal type and age (X21 = 0.26, *P* = 0.613), signal type and number of individuals in party (X21 = 2.15, *P* = 0.143), or signal type and sex (X21 = 2.47, *P* = 0.116) were found.

As rank was only available for older individuals, we ran a separate model to assess the effects of rank, by adding rank as a fixed effect and the interaction between rank and signal type to the full model specified above (*N* = 90 data points from 19 individuals). This version of the model that included rank (*N* = 19) confirmed that the overall model (X213 = 138.61, *P* < 0.001) and the Context\*UM/MM interaction (X23 = 19.51, *P* < 0.001) were significant; however, rank had no significant interaction with signal type (X21 = 0.08, *P* = 0.784).

# *Factors Affecting Unimodal Signal Production*

# As no individuals had sufficient time for UM facial expressions or any MM combination involving facial expressions available to calculate facial expression rates during the travel context, we ran two sets of models to examine (1) the effect of all contexts (including travel) on just vocal and gestural signals (facial expression excluded) and (2) the effect of a reduced set of context (excluding travel) on the full range of signals (facial expression included).

We first constructed a model to test whether variation in UM signal production rate (signals/h) could be explained by interactions between the following fixed factors: type of UM signal (gestures, vocalizations) and (1) context of production (rest, feed, groom, travel), (2) mean number of individuals in the party, (3) the age of the signaller and (4) the sex of the signaller. The full model comprised the above interaction terms and the associated fixed factors. Individual identity was included as a random factor. The dependent variable was rate of UM signal production/h. This model included the travel context, but excluded facial expressions. We included 184 data points from 26 individuals in the model.

Overall, the full model (*N* = 26 individuals) explained a significant amount of variation in signal production rates, compared to a null model (X213 = 82.24, *P* < 0.001). Likelihood ratio tests revealed that there was a significant interaction between type of UM signal and context (X23 = 57.87, *P* < 0.001; Fig. 3). Figure 3 illustrates that UM gestural signal production rates were significantly higher than UM vocal rates in rest and groom contexts. In contrast, in the travel context, UM vocal signal production rate was significantly higher than gestural production rates. In feed contexts, although vocalizations were given at higher rates than gestures, there was no significant difference between UM modality rates in this context. There were no significant interactions between UM signal type and age (X21 = 0.04, *P* = 0.843), UM signal type and the mean number of individuals in the party (X21 = 0.01, *P* = 0.917) or UM signal type and sex (X21 = 0.92, *P* = 0.338).

As rank was only available for older individuals, we ran a separate model to assess the effects of rank, by adding rank as a fixed effect and the interaction between rank and UM signal type to the full model specified above (*N* = 139 data points from 19 individuals).This version of the model that included rank (*N* = 19) confirmed the overall model (X215 = 71.82, *P* < 0.001), and the Context\*Modality of UM signal interaction (X23 = 46.52, *P* < 0.001) was significant; however, rank had no significant interaction with modality of the UM signal (X21 = 0.45 , *P* = 0.504).

Figure 3. The UM signal production rate (per h) of vocal (V) and gestural (G) signals in the contexts rest, feed, groom and travel. Error bars represent bootstrapped 95% confidence intervals, based on 1000 iterations. Error bars that do not overlap represent a significant difference. UM rate and confidence intervals derived from data of *N* = 25 (vocalizations and gestures rest), *N* = 26 (vocalizations and gestures feed), *N* = 23 (vocalizations and gestures groom), *N* = 20 (vocalizations travel), *N* = 16 (gestures travel).

Second, we constructed a model to test whether variation in UM signal production rate (signals/h) could be explained by interactions between the modality/type of UM signal (facial, vocal, gestural) and (1) context of production (rest, feed, groom), (2) mean number of individuals in the party, (3) the age of the signaller and (4) the sex of the signaller. This model excluded the travel context but included facial expressions. The full model comprised the above interaction terms and the associated fixed factors. Individual identity was included as a random factor. The dependent variable was rate of UM signal production/h. We included 191 data points from 26 individuals in the model.

Overall, the full model (*N* = 26 individuals) explained a significant amount of variation in signal production rates, compared to a null model (X217 = 144.98, *P* < 0.001). Likelihood ratio tests revealed that there was a significant interaction between type of UM signal and context (X24 = 56.84, *P* < 0.001; Fig. 4). Figure 4 reveals a similar pattern of results as Fig. 3, regarding vocalizations and gestures in rest, feed and groom contexts; however, it also illustrates that the rate of facial signal production was significantly below that for vocal and gestural signals in all three contexts. No significant interactions between UM signal type and age (X22 < 0.01, *P* = 0.998), UM signal type and number of individuals in the party (X22 = 1.05, *P* = 0.591), or UM signal type and sex (X22 = 2.78, *P* = 0.250) were found.

As rank was only available for older individuals, we ran a separate model to assess the effects of rank, by adding rank as a fixed effect and the interaction between rank and UM signal type to the full model specified above (*N* = 146 data points from 19 individuals). This model that included rank (*N* = 19) confirmed the overall model (X220 = 129.76, *P* < 0.001), and the Context\*Modality of UM signal interaction (X24 = 40.26, *P* < 0.001) remained significant; however, rank had no significant interaction with modality of the UM signal (X22 = 0.90, *P* = 0.638).

Figure 4. The UM signal production rate (per h) of facial (F), vocal (V) and gestural (G) signals in the contexts rest, feed and groom. Error bars represent bootstrapped 95% confidence intervals, based on 1000 iterations. Error bars that do not overlap represent a significant difference. UM rate and confidence intervals derived from data of *N* = 23 (facial rest, vocal and gestural groom), *N* = 25 (vocal and gestural rest), *N* = 26 (vocal and gestural feed). Missing bars occurred when the facial rate was zero.

# *Recipient Responses: MM Signal Versus UM Components*

# We focused on examining the responses to the grunt + gesture MM signal and matched UM component grunts and gestures. We constructed a model to test whether variation in whether or not the focal individual’s signal elicited a response from individuals within 5 m could be explained by (1) the signal type produced (UM vocal, UM gestural, MM vocal-gestural), (2) the number of individuals within 5 m (1 –2 or 3+) or (3) the rank of the signaller. The dependent variable was whether or not there had been any response (Yes/No), fixed factors were the type of signal, the signaller’s rank and individuals within 5 m. Identity of the signaller and signal number, which denoted which UM and MM signals were matched together, were included as random factors. There were 104 data points from seven individuals in the model.

Overall, the full model (*N* = 7 individuals) explained a significant amount of variation in whether or not the focal individual’s signal elicited a response from recipients within 5 m, compared to a null model (X24 = 37.12, *P* < 0.001). Likelihood ratio tests revealed that there was a significant main effect of signal type produced (X22 = 34.16, *P* < 0.001; Fig. 5). Figure 5 shows that UM vocal signals were significantly less likely to elicit a response from recipients than MM vocal-gestural signals or UM gestural signals. There was no significant difference in the proportion of MM vocal-gestural and UM gestural signals that elicited a response. A trend for lower ranking individuals to be more likely to receive a response than higher ranking individuals was found, but this effect was not significant (X21 = 2.85, *P* = 0.092), nor was the effect of the number of individuals within 5 m of the focal individual (X21 = 2.61, *P* = 0.106).

Figure 5. The mean proportion of focal individual MM vocal-gestural (VG), UM vocal (V) and UM gestural (G) signals that elicited a response from recipients within 5 m. Error bars represent bootstrapped 95% confidence intervals, based on 1000 iterations. Error bars that do not overlap represent a significant difference. Data based on *N* = 7 individuals.

*Do MM signals elicit the same responses as their UM components?*

We investigated whether the main responses elicited by a MM vocal-gestural signal matched those elicited by either of its UM components. Main responses were defined as the most active response that was the closest to the final behavioural outcome. For instance, if in response to a focal individual signal, a recipient looked at the signaller, approached and groomed, the main response was taken to be grooming.

Of the seven individuals for whom we compared MM signals and their UM components, Table 6 shows the four individuals from whom the MM signal elicited a response, and thus the responses to the UM components could be compared to the response to the MM signal (see Supplementary Table S4 for main responses elicited from all MM and matched UM signals, including those that did not elicit a response/were ignored). While MM signals from two female individuals elicited responses that matched the majority of responses to their UM gestural but not UM vocal signals (dominance of gestural response), one male individual elicited different responses to his MM signal than either of the components (emergence).

Table 6. Instances where the main response of the UM vocal signal and UM gestural signal matched the main response of the MM vocal-gestural signal for each individual

|  |  |  |
| --- | --- | --- |
| ID | Proportion (numbers) of UM vocalizations whose main responses match MM signal responses | Proportion (numbers) of UM gestures whose main responses match MM signal responses |
| AT (male) | 0.57 (4/7) | 0.13 (1/8) |
| PB (male) | 0.00 (0/2) | 0.00 (0/2) |
| PO (female) | 0.13 (1/8) | 0.8 (12/15) |
| WL (female) | 0.00 (0/19) | 0.82 (9/11) |

See Supplementary Table S4 for details of the type of responses elicited.Only individuals whose MM signals received a response were included in this table.

**DISCUSSION**

Although MM signals may not be as common as UM signals, this study has documented the production of 48 different free MM combinations. While 22 of 26 individuals produced at least one free MM signal, each broad type of MM signal combination was observed to be produced by at least nine of the 26 focal individuals. This suggests that the vast majority of individuals have the capacity and motivation to flexibly and simultaneously combine signals from different modalities, albeit rarely. In the future, a largescale, collaborative approach to document free MM signal production across individuals’ life spans and across study sites would shed valuable light on the mechanisms underpinning the production of these combinations (e.g. social learning, individual innovation, innately predetermined) and whether cultural variation exists in the type of free MM signal combinations commonly produced.

The number of UM signals produced per h was found to be more than 10 times higher than the number of MM signals in our study. This contrasts sharply with the two previous captive studies to compare proportions of UM and MM signals, which both found much higher relative rates of MM signals. Pollick and de Waal’s (2007) chimpanzee signals consisted of, 56% gestures, 22.5% facial/vocal signals and 21.6% MM combinations of the two. Similarly, Taglialatela et al. (2015) found that approximately half of the vocalizations recorded were accompanied by another communicative signal/behaviour. The relatively low levels of MM signals to UM signals, compared to these previous studies (see Supplementary Fig. S5 and Table S6), could be attributable to several factors. First, we identified MM signals as ones with temporal overlap, whereas previous studies considered signals or behaviours produced within 2 s (Taglialatela et al., 2015) or 10 s (Pollick & de Waal, 2007) of each other as MM signals. Second, we only considered vocal, gestural and facial signals, whereas Taglialatela et al. (2015) also included combinations of behaviours such as play or chase with vocalizations to be MM signals. Third, our study shows the importance of context in influencing the relative rate of UM and MM signals, whereas previous studies did not examine the same contexts as this study, nor did they specifically compare different contexts. For instance, Pollick & de Waal did not include rest, but importantly did include social play. Play is highly interactive, and it is common for individuals to show a range of MM signals in this context, such as play face, laughter, and various manual and nonmanual gestures simultaneously, so this could also explain the higher proportion of MM signals recorded. Finally, these previous two studies were conducted in captivity, where the social and physical environment may favour higher rates of MM signalling. In an enclosed area individuals are not normally able to express fission –fusion behaviour and this may mean that individuals need to use more sophisticated signals to negotiate tense social interactions, where in the wild they could simply leave the party, or seek a greater distance from certain individuals. Furthermore, in a captive environment visibility is generally much higher than in the dense tropical rainforest, meaning that MM combinations including visual signals are more likely to be successfully received. Investigating MM communication in wild savannah chimpanzees could be an interesting avenue for future research to explore whether the differences between the levels of MM signals produced in the wild and captivity seen so far is due to strategies learnt to cope with the limited space in captivity and interactions with humans, or in fact the level of visibility.

Our results partially supported our hypotheses that MM signals would be more likely to elicit a response than UM signals: the likelihood of a response was significantly higher with a MM grunt + gesture signal than a UM grunt signal, but similar to the matching UM gesture signal. This suggests that in the context of these specific signals, adding a vocal signal to a gesture does not change the likelihood of eliciting a response; in contrast, adding a gesture signal to a vocalization significantly improves the chances of eliciting a response. This supports findings from Pollick & de Waal’s (2007) study that indicated that MM signals of gestures combined with a vocalization or facial expression were no more effective at eliciting responses than gestures alone. Although adding vocalizations to gestures may not increase the likelihood of obtaining a response, it may help disambiguate the signaller’s intended meaning or convey more information than the UM signals in isolation. Indeed, in one individual the responses elicited to the MM signal were different to both the vocal and gestural components, indicating MM signals in chimpanzees have the potential to have emergent functions. Equally, it could be the case here that vocalizations are used as attention-getting signals alongside gestures (similar to Leavens, Russell & Hopkins, 2010), for example when the recipient does not have their visual attention directed towards the signaller. In this case the gesture might be the signal to which the signaller actually wants the recipient to attend.

Descriptively, when examining MM signals that elicited a recipient response, in two of the four individuals the type of response elicited by the MM signal was more likely to match the response elicited by the gestural than the vocal components. Viewed in the framework of Partan and Marler (2005), this indicates that for these individuals this particular free MM signal may be best characterized as a nonredundant combination that retains a dominant gestural response. Whether similar findings would be obtained if a different type of vocalization had been focused on remains unclear. The grunt vocalization investigated here might be relatively ambiguous as it is frequently produced in a variety of contexts. In contrast, many of the gestures that were highly successful in eliciting responses (e.g. present groom) were highly specific to a groom context and had a clear and measurable recipient response. As our repertoire illustrates, wild chimpanzees produce a large array of free MM signals and further research needs to systematically investigate the recipient responses to these and their matched UM component signals in order to understand the range of functions free MM signals have in this species.

In contrast to our predictions, we found that the proportion of signals that elicited a response was not dependent on the rank of the individual who produced the signal, nor the number of individuals who were within 5 m of this individual. It could be that the likelihood of a response may be more influenced by the rank difference or degree of friendship between signaller and recipient rather than the absolute rank of the signaller. We were not able to accurately calculate such relative dyadic measures, as for the majority of signals it was difficult to discern which individual was the recipient, and potentially there could have been several. In terms of the number of potential receivers, it could be the case that the majority of signals are in fact directed at a specific individual (e.g. Schel, Machanda et al., 2013), even if this might be difficult for human observers to detect, and thus the number of other individuals in the vicinity may not be an important predictor of a response. In the grunt + gesture MM signals that we investigated this is likely to be particularly true, as most signals occurred in a groom context, where the signals are likely to be directed at the grooming partner. In addition, we only considered grunts, which are an example of a ‘proximal’ vocalization that Taglialatela et al (2009) showed were more likely to be directed towards specific individuals, and to be processed differently by recipients, compared to ‘broadcast’ vocalizations, such as pant hoots.

Relative rates of vocal, gestural and facial signal production varied as a function of context. It was predicted that wild chimpanzees might tailor their signalling to the recipient’s attentional state, as has been shown in captivity (Leavens et al., 2010), and there were indications of this in this wild population. The signal production rate of UM gestures was found to be significantly higher than UM vocalizations in rest and groom contexts. This could be because in these contexts the focal individual might be more likely to have the visual attention of the recipient (especially when in close proximity, such as during grooming), whereas during feed or travel recipients are less likely to have the visual attention of others. Conversely, during travel individuals appear to produce significantly higher rates of vocal signals, which are more likely to be received not only by members of their own party, but also by more distantly located individuals. Facial expressions were only observed in a rest context; for feed and groom contexts the 11 and nine individuals that met the time criteria for calculation of a signalling rate had a facial signal production rate of zero. Thus UM facial expressions were recorded very rarely, and rates were significantly below those of vocal and gestural signals in rest, feed and groom contexts. Also note that we only coded salient facial expressions (see Table 3) and had we applied full FACS coding (Vick, Waller, Parr, Pasqualini & Bard, 2007) to our videos, subtler facial movements might have been captured. Nevertheless, in the dense forest environment facial expressions alone may be difficult for receivers to detect, and they may be more effective when combined with other signals. Indeed, Fig. 2 shows that facial expressions are produced in all contexts in which we could measure them in combination with other signals. This highlights the importance of adopting a holistic MM approach to studying animal communication: facial expressions in wild chimpanzees are most commonly produced as part of MM signals and extracting facial expressions and analysing them in isolation from the composite signal is likely to lead to incorrect understanding of signal function.

Contrary to our predictions, none of the individual factors we included in our models influenced the rate of UM and MM signal production. Age, sex and rank did not significantly interact with UM/MM signal rate. This indicates that learning to combine signals from different modalities and how to effectively use different types of signals may occur relatively early in development, before early adulthood. Previous research has shown that infant and juvenile chimpanzees may produce several different gestural signals in sequences as a ‘fail-safe’ strategy to elicit a response. In contrast, more mature individuals were found to produce fewer, but more successful signals (Hobaiter & Byrne, 2011b), and no differences in gesture signalling strategies were seen between subadults (10–14/15 years) and adults. In line with these findings, we found that age of the sub-adult and adult individuals we studied did not influence the relative proportion of MM and UM signals produced. As Hobaiter & Byrne (2011b) found juveniles and infants often used rapid fire gesture sequences, probably to encourage recipients to respond, future research should investigate whether infants and juveniles adopt a similar strategy with signal combinations and produce a higher proportion of MM signals than adults.

By examining multiple modalities and their combinations simultaneously we have revealed free MM combinations and flexible usage of different types of UM signals across contexts. Facial expressions were rarely produced in isolation and instead were more commonly combined with other signals: artificially extracting facial expressions from these composite signals could lead to misunderstanding of signal function. We advocate a MM approach to gain a full understanding not only of animal communication, but also of the evolutionary roots of human language. Human language is a multimodal communication system, with gestures and facial expressions accompanying and modifying the meaning of speech, and this study has shown that the ability and motivation to flexibly combine different signals are present in wild chimpanzees, and thus are likely to be present in our last common ancestor. Further research into the function of different free MM combinations may reveal the potential for generativity (emergent function, Partan & Marler, 2005) and social learning of MM signal combinations, which would have significant impact on our understanding of the evolution of these key facets of language.

In conclusion, our results reveal an impressive repertoire of free MM signals, but that these signals are used rarely compared to gestures and vocalizations in isolation. Interestingly, facial expressions are more commonly produced as part of MM signals than in isolation in several contexts. Systematic investigation of the MM grunt + gesture signal and the UM component signals revealed MM signals were more likely to elicit a response than UM vocal signals, but not UM gestural signals, and several potential functions for this specific type of MM signal were identified. The relative rates of UM vocal, gestural and facial signals varied across contexts, indicating flexible use of different signalling modalities across contexts. The flexibility in communicative signalling this study has revealed, by adopting a MM approach, may represent an important cognitive foundation from which our own complex multimodal communication system could have evolved.

**ACKNOWLEDGMENTS**

We are grateful to the directors of Kibale Chimpanzee Project for permitting and supporting us to carry out this research on the Kanyawara community of chimpanzees. We are also thankful to the KCP field manager Emily Otali and the KCP field assistants, Dan Akaruhanga, Seezi Atwijuze, Sunday John, Richard Karamagi, James Kyomuhendo, Francis Mugurusi, Solomon Musana and Wilberforce Tweheyo, for their valuable assistance and support in the field. We thank Simon Townsend for statistical advice, and two anonymous referees for their constructive comments. This project was funded by a Leakey Foundation General Grant to C.W., and we appreciate the permission of the Uganda National Council for Science and Technology, the President’s Office and the Uganda Wildlife Authority for us to carry out this study in Uganda.

**Supplementary Material**

Supplementary material associated with this article is available, in the online version, at doi

**REFERENCES**

Altmann, J. (1974). Oberservational study of behaviour: sampling methods. *Behaviour, 49(3)*, 227-267. <http://dx.doi.org/10.1163/156853974X00534>.

Bates, D., & Maechler, M. (2009). Dai B lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-28. http://cran.R-project.org/package=lme4.

Chapman, C. A., & Wrangham, R. W. (1993). Range use of the forest chimpanzees of Kibale: implications for the understanding of chimpanzee social organization. *American Journal of Primatology*, *31*(4), 263-273.

Cooper, B. G., & Goller, F. (2004). Multimodal signals: enhancement and constraint of song motor patterns by visual display. *Science*, *303*(5657), 544-546. http://dx.doi.org/10.1126/science.1091099.

 Crawley, M. J. (2002). *Statistical computing: an introduction to data analysis using S-Plus.* Chichester, U.K.: Wiley.

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. <http://dx.doi.org/10.1007/s10071-014-0827-z>.

de Luna, A. G., Hödl, W., & Amezquita, A. (2010). Colour, size and movement as visual subcomponents in multimodal communication by the frog Allobates femoralis. *Animal Behaviour*, *79*(3), 739-745. http://dx.doi.org/10.1016/j.anbehav.2009.12.031.

de Vries, H. A. N., Stevens, J. M., & Vervaecke, H. (2006). Measuring and testing the steepness of dominance hierarchies. *Animal Behaviour*, *71*(3), 585-592. http://dx.doi.org/10.1016/j.anbehav.2005.05.015.

Faraway, J. J. (2006). Extending the Linear Model with R: Generalized Linear. *Mixed Effects and Nonparametric Regression Models*. Boca Raton, FL: CRC Press.

Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. *Applied Psychological Measurement*, *5*(1), 105-112.

Georgiev, A. V., Russell, A. F., Thompson, M. E., Otali, E., Muller, M. N., & Wrangham, R. W. (2014). The foraging costs of mating effort in male chimpanzees (*Pan troglodytes schweinfurthii*). *International Journal of Primatology*, *35*(3-4), 725-745. http://dx.doi.org/10.1007/s10764-014-9788-y.

Gruber, T., & Zuberbühler, K. (2013). Vocal recruitment for joint travel in wild chimpanzees. *PLoS One*, *8*(9), e76073. <http://dx.doi.org/10.1371/journal.pone.0076073>.

Hayashi, M. (2007). A new notation system of object manipulation in the nesting-cup task for chimpanzees and humans. *Cortex*, *43*(3), 308-318. http://dx.[doi:10.1016/S0010-9452(08)70457-X](http://dx.doi.org/10.1016/S0010-9452(08)70457-X).

Hebets, E. A. (2008). Seismic signal dominance in the multimodal courtship display of the wolf spider Schizocosa stridulans Stratton 1991. *Behavioral Ecology*, *19*(6), 1250-1257. http://dx.doi.org/10.1093/beheco/arn080.

Hebets, E. A., & Papaj, D. R. (2005). Complex signal function: developing a framework of testable hypotheses. *Behavioral Ecology and Sociobiology*, *57*(3), 197-214. http://dx.doi.org/10.1007/s00265-004-0865-7.

Hebets, E. A., & Uetz, G. W. (1999). Female responses to isolated signals from multimodal male courtship displays in the wolf spider genus *Schizocosa* (Araneae: Lycosidae). *Animal Behaviour*, *57*(4), 865-872.

Higham, J. P., & Hebets, E. A. (2013). An introduction to multimodal communication. *Behavioral Ecology and Sociobiology*, *67*(9), 1381-1388. http://dx.doi.org/10.1007/s00265-013-1590-x.

Hobaiter, C., & Byrne, R. W. (2011a). The gestural repertoire of the wild chimpanzee. *Animal Cognition*, *14*(5), 745-767. http://dx.doi.org/10.1007/s10071-011-0409-2.

Hobaiter, C., & Byrne, R. W. (2011b). Serial gesturing by wild chimpanzees: its nature and function for communication. *Animal Cognition*, *14*(6), 827-838. http://dx.doi.org/10.1007/s10071-011-0416-3.

Hopkins, W. D., Taglialatela, J. P., & Leavens, D. A. (2011). Do chimpanzees have voluntary control of their facial expressions and vocalizations. In A. Villain, J-L. Schwartz, C. Abry & J. Vauclair (Eds), *Primate communication and human language: Vocalisation, gestures, imitation and deixis in humans and non-humans*, vol. *1*, pp. 71-88. Amsterdam, Netherlands: John Benjamins. http://dx.doi.org/10.1075/ais.1.05hop.

Kalan, A. K., & Boesch, C. (2015). Audience effects in chimpanzee food calls and their potential for recruiting others. *Behavioral Ecology and Sociobiology*, *69*(10), 1701-1712. http://dx.doi.org/10.1007/s00265-015-1982-1.

Leavens, D. A., Hopkins, W. D., & Bard, K. A. (1996). Indexical and referential pointing in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, *110*(4), 346.

Leavens, D. A., Hopkins, W. D., & Thomas, R. K. (2004). Referential communication by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, *118*(1), 48. http://dx.doi.org/10.1037/0735-7036.118.1.48.

Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2010). Multimodal communication by captive chimpanzees (*Pan troglodytes*). *Animal Cognition*, *13*(1), 33-40. http://dx.doi.org/10.1007/s10071-009-0242-z.

Liebal, K., Waller, B. M., Slocombe, K. E., & Burrows, A. M. (2013). *Primate Communication: a multimodal approach*. Cambridge, U.K.: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139018111 .

Muller, M. N., & Wrangham, R. W. (2014). Mortality rates among Kanyawara chimpanzees. *Journal of Human Evolution*, *66*, 107-114.  [http://dx.doi.org/10.1016/j.jhevol.2013.10.004](%20http://dx.doi.org/10.1016/j.jhevol.2013.10.004)

Nishida, T., Kano, T., Goodall, J., McGrew, W. C., & Nakamura, M. (1999). Ethogram and Ethnography of Mahale Chimpanzees. *Anthropological Science*, *107*(2), 141-188. http://dx.doi.org/10.1537/ase.107.141.

Parr, L. A. (2004). Perceptual biases for multimodal cues in chimpanzee (*Pan troglodytes)* affect recognition. *Animal Cognition*, *7*(3), 171-178. http://dx.doi.org/10.1007/s10071-004-0207-1.

Parr, L. A., Waller, B. M., Vick, S. J., & Bard, K. A. (2007). Classifying chimpanzee facial expressions using muscle action. *Emotion*, *7*(1), 172. http://dx.doi.org/10.1037/1528-3542.7.1.172.

Partan, S. R. (2002). Single and multichannel signal composition: facial expressions and vocalizations of rhesus macaques (*Macaca mulatta*). *Behaviour*, *139*(8), 993-1027. http://dx.doi.org/[10.1163/15685390260337877](http://dx.doi.org/10.1163/15685390260337877).

Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science*, *283*, 1272-1273.

Partan, S. R., & Marler, P. (2005). Issues in the classification of multimodal communication signals. *The American Naturalist*, *166*(2), 231-245. http://dx.doi.org/10.1086/431246.

Partan, S. R., Fulmer, A. G., Gounard, M. A., & Redmond, J. E. (2010). Multimodal alarm behavior in urban and rural gray squirrels studied by means of observation and a mechanical robot. *Current Zoology*, *56*(3), 313-326.

Partan, S. R., Larco, C. P., & Owens, M. J. (2009). Wild tree squirrels respond with multisensory enhancement to conspecific robot alarm behaviour. *Animal Behaviour*, *77*(5), 1127-1135. http://dx.doi.org/10.1016/j.anbehav.2008.12.029.

Pollick, A. S., & De Waal, F. B. (2007). Ape gestures and language evolution. *Proceedings of the National Academy of Sciences*, *104*(19), 8184-8189. http://dx.doi.org/10.1073/pnas.0702624104.

Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. *Animal Behaviour*, *58*(5), 921-931. [http://dx.doi.org/10.1006/anbe.1999.1242](http://psycnet.apa.org/doi/10.1006/anbe.1999.1242).

Schel, A. M., Machanda, Z., Townsend, S. W., Zuberbühler, K., & Slocombe, K. E. (2013). Chimpanzee food calls are directed at specific individuals. *Animal Behaviour*, *86*(5), 955-965.  [http://dx.doi.org/10.1016/j.anbehav.2013.08.013](%20http://dx.doi.org/10.1016/j.anbehav.2013.08.013).

Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., & Slocombe, K. E. (2013). Chimpanzee alarm call production meets key criteria for intentionality. *PLoS One*, *8*(10), e76674. <http://dx.doi.org/10.1371/journal.pone.0076674>.

Schwarz, A. (1974). Sound production and associated behaviour in a cichlid fish, *Cichlasoma centrarchus*. *Zeitschrift fuer Tierpsychologie*, *35*(2), 147-156.

Slocombe, K. E., Kaller, T., Turman, L., Townsend, S. W., Papworth, S., Squibbs, P., & Zuberbühler, K. (2010). Production of food-associated calls in wild male chimpanzees is dependent on the composition of the audience. *Behavioral Ecology and Sociobiology*, *64*(12), 1959-1966. http://dx.doi.org/10.1007/s00265-010-1006-0.

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. http://dx.doi.org/10.1016/j.anbehav.2011.02.002.

Slocombe, K. E., & Zuberbühler, K. (2005). Functionally referential communication in a chimpanzee. *Current Biology*, *15*(19), 1779-1784.  [http://dx.doi.org/10.1016/j.cub.2005.08.068](%20http://dx.doi.org/10.1016/j.cub.2005.08.068).

Slocombe, K. E., & Zuberbühler, K. (2006). Food-associated calls in chimpanzees: responses to food types or food preferences? *Animal Behaviour*, *72*(5), 989-999. http://dx.doi.org/10.1016/j.anbehav.2006.01.030.

Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences*, *104*(43), 17228-17233. http://dx.doi.org/10.1073/pnas.0706741104.

Slocombe, K. E., & Zuberbühler, K. (2010). Vocal communication in chimpanzees. In [E. V. Lonsdorf](http://press.uchicago.edu/ucp/books/author/L/E/au8331039.html), [S. R. Ross](http://press.uchicago.edu/ucp/books/author/R/S/au8361706.html) & [T. Matsuzawa](http://press.uchicago.edu/ucp/books/author/M/T/au8361707.html) (Eds.) *The mind of the chimpanzee: ecological and experimental perspectives* (pp. 192-207)*.* Chicago, IL: University of Chicago Press.

Smith, W. J. (1977). The behavior of communicating: an ethological approach. Harvard University Press, Cambridge, MA.

Smith, C. L., & Evans, C. S. (2013). A new heuristic for capturing the complexity of multimodal signals. *Behavioral Ecology and Sociobiology*, *67*(9), 1389-1398. http://dx.doi.org/10.1007/s00265-013-1490-0.

Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2009). Visualizing vocal perception in the chimpanzee brain. *Cerebral Cortex*, *19*(5), 1151-1157. <http://dx.doi.org/10.1093/cercor/bhn157>.

Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2011). Chimpanzee vocal signaling points to a multimodal origin of human language. *PloS one*, *6*(4), e18852. http://dx.doi.org/10.1371/journal.pone.0018852.

Taglialatela, J. P., Russell, J. L., Pope, S. M., Morton, T., Bogart, S., Reamer, L. A., Schapiro, S. J., & Hopkins, W. D. (2015). Multimodal communication in chimpanzees. *American journal of primatology*, *77*(11), 1143-1148. http://dx.doi.org/10.1002/ajp.22449.

Tomasello, M. (2008). *The origins of human communication*. Cambridge, MA: MIT Press.

Townsend, S. W., & Züberbuhler, K. (2009). Audience effects in chimpanzee copulation calls. *Communicative & integrative biology*, *2*(3), 282-284. http://dx.doi.org/10.4161/cib.2.3.6796.

Uetz, G. W., & Roberts, J. A. (2002). Multisensory cues and multimodal communication in spiders: insights from video/audio playback studies. *Brain, Behavior and Evolution*, *59*(4), 222-230. [http://dx.doi.org/10.1159/000064909](http://psycnet.apa.org/doi/10.1159/000064909).

Uetz, G. W., Roberts, J. A., & Taylor, P. W. (2009). Multimodal communication and mate choice in wolf spiders: female response to multimodal versus unimodal signals. *Animal Behaviour*, *78*(2), 299-305. http://dx.doi.org/[10.1016/j.anbehav.2009.04.023](http://dx.doi.org/10.1016/j.anbehav.2009.04.023).

Waller, B. M., Liebal, K., Burrows, A. M., & Slocombe, K. E. (2013). How can a multimodal approach to primate communication help us understand the evolution of communication? *Evolutionary Psychology*, *11*(3), 147470491301100305. <http://dx.doi.org/10.1177/147470491301100305>.

Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., ... & Slocombe, K. E. (2015). Vocal learning in the functionally referential food grunts of chimpanzees. *Current Biology*, *25*(4), 495-499. http://dx.[doi:10.1016/j.cub.2014.12.032](http://dx.doi.org/10.1016/j.cub.2014.12.032).

Vick, S. J., Waller, B. M., Parr, L. A., Pasqualini, M. C. S., & 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. <http://dx.doi.org/10.1007/s10919-006-0017-z>.

Wilson, M. L., Kahlenberg, S. M., Wells, M., & Wrangham, R. W. (2012). Ecological and social factors affect the occurrence and outcomes of intergroup encounters in chimpanzees. *Animal Behaviour*, *83*(1), 277-291. http://dx.[doi:10.1016/j.anbehav.2011.11.004](http://dx.doi.org/10.1016/j.anbehav.2011.11.004).

Wrangham, R. W., Clark, A. P., & Isabirye-Basuta, G. (1992). Female social relationships and social organization of Kibale Forest chimpanzees. *Topics in primatology*, *1*, 81-98.