1 **Production of and responses to unimodal and multimodal signals in wild**

2 **chimpanzees, *Pan troglodytes schweinfurthii***

3 Claudia Wilke[a], Eithne Kavanagh[a], Ed Donnellan[a,b], Bridget M. Waller[c], Zarin P.

4 Machanda[d,e], Katie E. Slocombe[a]

5

6 [a] Department of Psychology, University of York, York, U.K.

7 [b] Department of Psychology, University of Sheffield, Sheffield, U.K.

8 [c] Department of Psychology, University of Portsmouth, Portsmouth, U.K.

9 [d] Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, U.S.A.

10 [e] Department of Anthropology, Tufts University, Medford, MA, U.S.A.

15 **Correspondence**: K. Slocombe, Department of Psychology (office B201), University of

16 York, Heslington, York YO10 5DD, U.K.

17 E-mail address: Katie.slocombe@york.ac.uk

18

19

20

21

22

23

24

25

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

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).

101   In line with a framework offered by Partan and Marler (1999), the function of a MM signal

102   can be determined by comparing recipient responses to the MM signal and the UM

103   components in isolation. In the case of fixed vocal-visual MM signals, this has often been

104   determined through careful experiments that used playbacks for vocal signals and animated

105   models to test responses to visual signals. Although experiments remain the best way to study

106   MM signal function and have been applied to free MM signals (Partan, Larco & Owens,

107   2009; 2010), the function of these signals can also be examined by collection of careful

108   observational data on recipient responses to the MM signal and its component parts when

109   produced unimodally. Broadly, MM signals can be categorized into (1) redundant

110   combinations where recipients produce the same response to the component UM signals and

111   the MM signal, but the response to the MM signal may be enhanced, and (2) nonredundant

112   combinations where recipients produce different responses to the component UM signals,

113   with possibilities for the responses to the MM signal to be different from those to the UM

114   components (emergence) or more similar to those to one of the UM signals (dominance). To

115   date, although MM signals are well documented in the animal kingdom, and have been

116   rigorously investigated with elegant experiments in a number of nonprimate species, there is

117   a lack of comparable investigation into MM communication in primate species (Liebal et al.

118   2013).

119   Understanding the communicative abilities of primates is not only important for establishing

120   a window into their complex social world and cognitive abilities, but also for understanding

121   human language evolution. Mapping out the differences and similarities in communicative

122   abilities of humans and our closest living relatives may help us discern which are the derived,

123   uniquely human aspects of language and which may have built on abilities already present in

124   common ancestors with extant primates. In addition, characteristics of primate vocal and

125   gestural communication provide key lines of evidence for theories concerning whether

126   language has vocal or gestural origins (Slocombe, Waller & Liebal, 2011). Among the

127   primates, chimpanzees, our closest living relatives, provide the best model of what our last

128   common ancestor might have been capable of, and thus play a critical role in informing

129   debates on the evolutionary origins of human language (Hayashi, 2007; Watson et al, 2015;

130   Schel, Machanda, Townsend, Zuberbühler & Slocombe, 2013; Taglialatela, Russell,

131   Schaeffer & Hopkins, 2011; Hobaiter & Byrne, 2011a).

132   For chimpanzees, vocalizations, gestures and facial expressions are all important

133   forms of communication, and previous UM research on these different types of signals have

134   investigated characteristics such as intentionality (e.g. Leavens, Hopkins & Thomas, 2004;

135   Schel, Townsend, Machanda, Zuberbühler, & Slocombe, 2013; Hopkins, Taglialatela &

136   Leavens, 2011), referentiality (e.g. Slocombe & Zuberbühler, 2005; 2006; Crockford, Wittig

137   & Zuberbühler, 2015), flexible use across contexts (e.g. Hobaiter & Byrne, 2011) and

138   audience effects (e.g. Gruber & Zuberbühler, 2013; Leavens, Hopkins & Bard, 1996; Kalan

139   & Boesch, 2015; Schel, Machanda, et al, 2013; Slocombe & Zuberbühler, 2007; Slocombe et

140   al. 2010; Townsend & Zuberbühler, 2009). On the surface, this UM work indicates that

141   gestures, vocalizations and facial expressions differ in terms of these characteristics;

142   however, few studies have attempted to examine these characteristics in a comparable

143   manner in multiple modalities, so such conclusions may be premature (Slocombe et al.,

144   2011). One study that has successfully examined different types of signal within a single

145   experimental paradigm explored whether captive chimpanzees could selectively produce a

146   signal appropriate to the attentional state of a human. Leavens, Russell and Hopkins (2010)

147   showed that chimpanzees, while begging from a human experimenter, used more visual

148   gestural signals when the researcher was facing towards them, and more tactile and vocal

149   signals when they were facing away.

150   Despite the wealth of research on the production of vocal, gestural and facial signals in

151   isolation, the combination of these signal types into MM signals in chimpanzees is virtually

152   unexplored (Liebal et al., 2013; Slocombe, Waller & Liebal, 2011). Important exceptions to

153   this include an experimental study probing recipient integration of signals from different

154   modalities, which revealed that chimpanzees can cross-modally match facial expressions and

155   vocalizations (Parr, 2004). In addition, Parr found that either the vocal or facial components

156   were more salient to the chimpanzees depending on the signal type (e.g. the vocal component

157   of a pant hoot signal was more salient than the facial component). From a production

158   perspective, a recent study by Taglialatela et al. (2015) indicated that approximately 50% of

159   captive chimpanzee vocalizations were accompanied by nonvocal signals (e.g. gestures, fear

160   grimace) or behaviours (e.g. chase, play), and that these combined signals were more likely to

161   be directed towards another individual than vocal signals alone. This indicates that

162   chimpanzees may use signal combinations from different modalities strategically to meet

163   specific sociocommunicative goals. Focusing on the combination of gestural signals with

164   vocal or facial signals in captive chimpanzees, Pollick and de Waal (2007) found 21% of

165   chimpanzee signals were MM. However, the operational definition of MM signals probably

166   captured MM sequences as well as signals, as signals occurring within 10 s of each other

167   were counted as MM signals. Perhaps surprisingly, MM signals were not found to be more

168   effective in eliciting a response than UM signals. However, unfortunately, this study's

169   findings are difficult to interpret as the analyses also suffer from pseudoreplication (Waller et

170   al. 2013). Despite variation in how these two studies define a MM signal, it seems that in

171   captivity, where visibility of group members is usually excellent, vocal, gestural and facial

172   signals may be commonly combined into MM sequences or signals. The degree to which

173   chimpanzees produce MM signals in their visually dense natural habitat, and whether in a

174   wild setting MM signals are more effective at eliciting responses than UM signals, remains

175   unknown. In addition, despite free MM signals having the potential to generate new meaning

176   (emergent function; Partan & Marler, 1999) and to indicate cognitive complexity relevant to a

177   language evolution perspective (Slocombe, Waller & Liebal, 2011), we are currently lacking

178   a MM repertoire and an understanding of how common and varied such free combinations

179   may be.

180   In this study we attempted to address these issues and systematically investigated the UM and

181   MM communication of wild chimpanzees, by taking an integrated MM approach. We

182   considered MM signals temporally overlapping combinations of vocal, gestural and facial

183   signals. We aimed to provide the first MM signal repertoire, understand the individual and

184   contextual factors that affect UM and MM signal production, and compare the recipient

185   responses to MM and matched UM signals.

186   In terms of signal production, we predicted that the rate of UM signal production would be

187   significantly higher than that of MM signals. Furthermore, we expected MM rates may be

188   lower than those found in captivity, due to adoption of stricter criteria and the more restricted

189   transmission of visual signals in a dense forest environment. Second, in terms of individual

190   factors, we expected that younger, female or more subordinate individuals may show higher

191   rates of MM than UM signals, as they may need to show more elaboration in signalling in

192   order to elicit responses than older, male, dominant individuals. Third, focusing on UM

193   signals, given that captive chimpanzees modulate signal type depending on the recipient's

194   visual attention (Leavens, Russell & Hopkins, 2010), we predicted that relative rates of vocal,

195   gestural and facial signals would vary with context, with higher rates of gestures and facial

196   expressions in contexts where visual signals would be most visible for receivers (e.g. rest,

197   groom).

198 In terms of recipient responses, we predicted that MM signals would elicit significantly

199 higher proportions of recipient responses than matched UM signals, as increased recipient

200 responses to MM signals have been repeatedly found in rigorous nonprimate studies

201 (reviewed in Liebal et al., 2013), because recipients are more likely to detect and attend to

202 these more elaborate and salient signals. We also predicted that recipient responses would be

203 more likely when the signaller was more dominant and there were more recipients in the

204 vicinity.

205

206 **METHODS**

207

208 *Study Site and Subjects*

209 This study was carried out in Kibale National Park, located in western Uganda (0"13' –

210 0"41'N and 30"19' –30"32'E) in 2013 –2015. A detailed description of the characteristics of

211 the forest can be found in Chapman and Wrangham (1993). The study animals were a wild

212 group of chimpanzees, the Kanyawara community. In 2013, the group comprised

213 approximately 57 individuals (Muller & Wrangham, 2014), and occupied a home range of

214 around 16.4 km² (Wilson, Kahlenberg, Wells & Wrangham, 2012). The community is

215 entirely habituated and have been followed and studied regularly since 1987 by the Kibale

216 Chimpanzee Project (Wrangham, Clark & Isabirye-Basuta, 1992; Georgiev et al., 2014).

217 Specifically, the individuals included in this study were 13 males and 13 females, from 8 to

218 47 years old (see Table 1). These individuals were chosen on the basis that they were easy to

219 find and follow, ensuring that as much high-quality focal time as possible could be collected

220 for each individual. Dominance ranks were established by calculating a modified David's

221 score, MDS (de Vries, Stevens & Vervaecke, 2006), for all individuals for which long-term

222 field assistant data on decided aggressive interactions and submissive pant grunt

223 vocalizations were available (these data were unavailable for some younger individuals; their

224 rank was noted as NA). MDS was calculated for males and females separately and all males

225 were ranked above all females, as all of these males had dominated the females.

226

227 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

228 Age in 2013, the first year of data collection. Rank order is based on the modified David's

229 score. NA indicates young individuals for whom these data were not available.

230

231 *Equipment*

232 All focal observational data were collected with a Panasonic HDC-SD90 camcorder, with a

233 Sennheiser MKE 400 microphone attached. Recipient responses were recorded with a second

234 camcorder: a Panasonic HDC-SD40. Videos were coded using Noldus Observer XT 10 event

235 logging software (http://www.noldus.com/animal-behavior-research) for observational data.

236

237 *Ethical Note*

238 This study complied with the ASAB/ABS guidelines for the use of animals in research;

239 ethical approval was granted by the Biology Ethics Committee (University of York). The

240 Ugandan Wildlife Authority and the Ugandan National Council for Science and Technology

241 granted permission to carry out the study in Uganda.

242

243 *Data Collection*

244 All data were collected February–May 2013 and June 2014 –March 2015, between 0800 and

245 1830 hours. Focal animal sampling (Altmann, 1974) was employed in order to collect

246 observational data on the 26 focal individuals. Focal animals were only sampled once a day

247 and were chosen in a way that maximized the quality and spread of data across target

248 individuals. Initially, once a party of chimpanzees were located, the target animal with the

249 best visibility for clear filming was chosen as the focal individual, but later on in the study

250 period target individuals with the least focal time were prioritized.

251

252 **Focal animal signal production**

253 Focal samples consisted of 15 min of continuous video observation of one focal animal. The

254 aim was to capture on video a complete view of all facial, gestural and vocal signals

255 produced by the focal individual. Thus the camera was zoomed in as close as possible, while

256 still capturing the whole body of the chimpanzee. The researcher commentated all

257 vocalizations in real time, to ensure that even quiet vocalizations that could not be picked up

258 by the microphone were recorded. If individuals moved out of sight and earshot during a

259 focal observation, this time was coded as 'out of sight' (OOS) and excluded from any further

260 analysis. Samples containing more than 10 min of OOS time were excluded from further

261 analysis, meaning the analysed samples range from 5 to 15 min and all had a good level of

262 visibility of the focal animal. As we were interested in social communication, only focal

263 samples during which the focal individual was in a party (i.e. there were other independent

264 individuals within 30 m; Slocombe et al, 2010) were included in this analysis. Thus this

265 excluded times when mothers were alone, with only their infants, as mother – infant

266   communication could not be examined in the majority of the focal individuals, who were not

267   mothers.

268

269   **Recipient responses**

270   To collect the response of other individuals to any signals produced by the focal individual, a

271   second researcher used a camcorder to capture the signals and behaviour of as many of the

272   individuals closest to the focal animal (within 5 m) as possible.

273

274   *Video Coding*

275   Video coding with Observer XT 10 software was used to extract continuous details about the

276   behavioural context and modality availability of the focal individual, as well as all their UM

277   and MM signal production (see detailed definitions below). By coding the context and

278   modality availability continuously we were able to calculate accurate rates of signal

279   production in each context, as a function of the time each specific modality could be reliably

280   detected. The types of behaviours elicited from individuals within 5 m of the focal individual,

281   in response to each focal signal, were also coded from the videos (see detailed definitions of

282   these responses below). Recipient responses were only coded for the 32 h of video data for

283   which a second observer was present to capture these on a second video camera.

284

285   *Definitions*

286   **Behavioural contexts**

287 Eight behavioural contexts were defined and coded, but only four occurred frequently enough

288 across focal animals to be examined further in terms of signal production rates (rest, groom,

289 feed, travel; see Table 2). For these four contexts, the requisite behaviour had to continue for

290 at least 20 s (a break of up to 5 s is permissible during this time), after which breaks of up to

291 15 s were permissible, as long as the individual always returned to the original behaviour. For

292 the repertoire and number of instances of different MM combinations (see Table 5 below and

293 Table S3 in the Supplementary material), all contexts were included i.e. all available video

294 time, in order to establish a more comprehensive picture of the types of signals that the focal

295 individuals were motivated to combine.

296

297 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)

298

299   Some behaviours could occur simultaneously; thus there was a hierarchy when coding, with

300   the more active (generally also rarer) behaviour being given precedence: travel>feed,

301   display>travel, play>travel, aggression>travel, aggression>display, play>feed, groom with

302   other>self-groom. It was occasionally the case that an individual would be involved in an

303   agonistic interaction and then rest, feed, travel or groom immediately afterwards. In these

304   cases, behavioural contexts were still based on the current behaviour of the individual; thus

305   rest, feed, groom and travel contexts also include postconflict periods.

306

307   **Modalities available**

308   This behavioural coding group was used to capture which type of signals produced by the

309   focal individual could be coded reliably from the video at all times (see Supplementary Table

310   S1). It was frequently the case that only signals in one or two modalities could be accurately

311   captured due to the orientation of the focal animal (e.g. face may not be visible) or distance of

312   observer to the chimpanzee (e.g. quiet vocalizations may not be detected). This was coded so

313   it could be taken into account when calculating rates of signal production.

314   To be coded as a 'modality available', the requisite modality had to be available for at least

315   20 s (a break of up to 5 s is permissible during this time), after which breaks of up to 15 s

316   were permissible, as long as the original modality then became available again. The

317   exception to this rule was when the modality could not be seen for most of the time, but in the

318   short period for which it was available, a signal was produced (for instance an individual's

319   face cannot be seen, they then turn around for 2 s, showing a 'play face', then turn away

320   again); in this case it was coded as available for this short period, and the signal produced

321   was also coded.

322

323   **Focal animal signals produced**

324   We coded all vocal, gestural (manual and nonmanual) and facial signals the focal individual

325   produced (see Table 3). The duration of facial and gestural signals was coded; for

326   vocalizations, which were commonly produced in bouts, the duration of the calling bout was

327   recorded. Two or more vocalizations of the same type were coded as one continuous bout if

328   they were produced within 10 s of one another (from the end of one to the beginning of the

329   next). Eight different facial expressions were coded; these were based on the prototypical

330   expressions described in Parr, Waller, Vick and Bard (2007), which are specific combinations

331   of facial muscle movements (Action Units: ChimpFACS, Vick et al. 2007). The person

332   coding the signals discussed exemplars with B.W. (certified FACS coder) prior to video

333   coding in order to avoid any expressions that did not fit the prototypical descriptions. Forty

334 common gestures were coded based on the repertoire proposed by Hobaiter & Byrne (2011a).

335 Rare gestures were coded as 'other manual gesture' or 'other nonmanual gesture' and

336 described in the notes section. Similarly, owing to the size limits of the coding scheme, some

337 gestures were combined under an umbrella term, for instance 'manual contact with another

338 individual' included touch, hand on, punch, push, slap, tap, poke, hit. Fourteen different

339 vocalizations were coded based on the repertoire proposed by Slocombe and Zuberbühler

340 (2010).

341

342 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

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

344

345 **Recipient response time and types**

346 Recipient responses were coded from the beginning of the focal individual signal until 20 s

347 after the signal had finished, from individuals within 5 m of the focal individual. During the

348 recipient response time the number and identity of the individuals within 5 m of the focal

349 individual were recorded. If another signal occurred within the 20 s after the first signal then

350 the recipient response time was cut short for the first, with this only lasting until the

351 beginning of the next signal. Similarly, if the recipient response continued after the 20 s (for

352 example the signal elicited a long bout of grooming), this was also only coded for up to 20 s

353 after the end of the focal signal.

354 Recipient responses comprised four groups: signal responses (facial, vocal, gestural and

355 MM), movements, negative and positive responses (see Table 4). Signals by other individuals

356 were only coded as responses if the recipient's facial expression or gesture was directed at the

357 focal individual (as far as this was relevant and possible to discern). It was difficult to

358 determine specifically to whom vocalizations were directed, so all vocalizations from

359 recipients were counted as potential signals in response to the focal individual. Any signals or

360 behaviours that were clearly in response to an unrelated signal or event were not coded as

361 responses. For example, if the focal individual gave a big loud scratch (BLS) gesture, and

362 immediately afterwards individuals in another party uttered pant hoots and an individual

363 within 5 m of the focal animal replied with a pant hoot, the pant hoot was not coded as a

364 response to the BLS. Similarly, only an active change in behaviour of the recipient was coded

365 as a response. For instance, if another individual was already vocalizing, and then the focal

366 individual produced a signal, and the other individual continued vocalizing as before, this

367 was not counted as a response to the focal animal's signal. Equally, 'terminating' behaviours

368 were not coded, for example the cessation of playing or grooming. Behavioural responses

369 (positive, negative) had to be directed towards the focal animal rather than a third party to be

370 counted as a response to the focal animal's signal.

371

372 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)

373

374

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.

389

390   Rates were only calculated for MM combinations for a specific context for an individual if

391   they had at least 15 min of this modality combination available in rest, feed and groom

392   contexts, and at least 10 min in the travel context. For example, to contribute a rate for vocal-

393   gestural signals in the travel context, that individual must have had at least 10 min of time

394   where both vocalizations and gestures were simultaneously available in this context. Mean

395   MM rate was the average of facial-gestural, vocal-gestural, fixed facial-vocal, free facial-

396   vocal and facial-vocal-gestural rates. Individuals must have contributed at least three of these

397   MM combination rates to have a mean MM rate calculated in a specific context. Those who

398   did not meet this had a missing value for MM rate in this context.

399   Mean signal production rates for the group (as reported in descriptive statistics and figures in

400   the Results) were calculated as a mean of all the individual mean production rates that

401   contributed to a particular analysis.

402   To assess whether the number of individuals in the party affected signal production, we

403   calculated the average number of individuals in the party present during the periods from

404   which signal rates were calculated for each type of signal produced by each individual. First,

405   the number of individuals in the party was recorded at the beginning and end of every video

406   and these were averaged. Second, for each signal type for which a rate was calculated for an

407   individual, we took the corresponding videos that had contributed to the calculation of that

408   rate and calculated a mean from the average number of individuals in the party across those

409   videos.

410

411   *Comparison of responses to MM signals and UM components*

412   Most previous nonprimate research carried out within a MM framework has focused on fixed

413   MM signals and/or signals produced only in one specific context, for example alarm

414   behaviour (e.g. Partan et al., 2009) or courtship behaviour (e.g. Uetz, Roberts & Taylor,

415   2009). In contrast, the signals we investigated were free MM signals, which were produced

416   across a range of contexts (see Supplementary Table S4). As context was shown to heavily

417   influence signal production (see signal production results below), we endeavoured to control

418   for this by matching UM and MM signals based on signaller identity and behavioural context

419   of production. We consider such matching of MM signals and UM component signals to be

420   critically important in order to understand the function of the signals.

421

422   We focused on the free MM signal produced most frequently by the largest number of

423   individuals, where matched UM component signals were also frequently produced by the

424   same individuals: the grunt + gesture signal (vocal-gestural combination). It was not possible

425   to examine more MM signal combinations as no other type of free MM signal, with sufficient

426   matched UM components, was produced by a sufficient number of individuals.

427   For each of the MM signals we identified component UM signals that were matched to the

428   MM signal in terms of the behavioural context during production. Up to five UM grunt

429   signals and five UM gesture signals were matched to each MM signal. Where possible we

430   also matched the number of individuals present within 5 m of the focal individual. For

431   instance, if the individual PO produced a grunt + present groom MM signal in a groom

432   context, with two individuals within 5 m, the responses to this signal could be compared to

433   the responses to a UM grunt vocalization from PO, in a groom context, with two individuals

434   within 5 m of her, and a UM present groom gesture, in a groom context, with three

435   individuals within 5 m of her.

436

Intercoder Reliability

438   To assess the intercoder reliability of the video coding, a second independent researcher also

439   coded 6.5% of the videos (7.75 h, $N = 31$ videos each lasting 15 min from a total of 15

440   individuals), having been provided with comprehensive instructions. Cohen's kappa was

441   calculated; the mean Kappa value obtained was 0.81, indicating excellent levels of coder

442   agreement (Fleiss, 1981). All reliability analyses were run using the Reliability Analysis

443   function in Observer XT 10, which enables the comparison of two different Event Logs for

444   one video.

445

446   *Data Analysis*

447   We constructed linear mixed models (LMMs) and generalized linear mixed-effects models

448   (GLMMs) in order to test our hypotheses regarding signal production and recipient responses

449   respectively. LMMs were used to investigate the influence of continuous and categorical

450   variables on signal production rates, while GLMMs with a binomial error structure were used

451   to investigate the influence of continuous and categorical variables on the occurrence of

452   recipient responses (binary response variable: received one or more responses or no

453   responses). Furthermore, because we had repeated sampling from the same individual, to

454   control for pseudoreplication we fitted 'individual' as a random factor (Crawley, 2002) by

455   conducting random intercepts models using the package lme4 (Bates & Maechler, 2009;

456   https://cran.r-project.org/web/packages/lme4/index.html). We first assessed whether the full

457   model could explain a significant amount of variation in the dependent variable, by

458   comparing the full model to a null model containing just the intercept and random factors. To

459   assess the significance of each explanatory variable or interaction term, we compared the full

460   model with a reduced model excluding the variable or interaction of interest using a

461 likelihood ratio test (Faraway, 2006). All models were run in R v. 2.15 (The R Foundation for

462 Statistical Computing, Vienna, Austria, http://www.r-project.org).

463

464 As some data were not available for all individuals (e.g. dominance rank) or were missing in

465 the majority of individuals due to methodology (e.g. facial expression was not available

466 during travel, as the observer followed and filmed travelling chimpanzees from behind), we

467 sometimes constructed several models in order to test our hypotheses thoroughly, and to

468 maintain a high number of individuals in each model.

469

470 **RESULTS**

471

472 *MM Signals: Repertoire, Rates and Responses*

473 Overall, the results show that across rest, feed and groom contexts MM signals were rare

474 relative to UM signals (see Fig. 1). Free MM signals were, however, produced by 22 of the

475 26 focal individuals, and we recorded a total of 48 different free MM signals, consisting of

476 combinations that in total included six different facial expressions, nine different

477 vocalizations and 16 different gestures (see detailed MM repertoire in Supplementary Table

478 S3). Vocal-gestural combinations were the most common free MM signals recorded, and free

479 facial-vocal the least (see Table 5). The frequency of different types of responses the various

480 different categories of MM signal elicited from those within 5 m are also shown in Table 5.

481 Vocal-gestural signals were the most likely to elicit any kind of response, and the most likely

482 of all the signal combinations to elicit a positive response. In contrast fixed facial-vocal

483 signals received the highest percentage of negative responses.

484
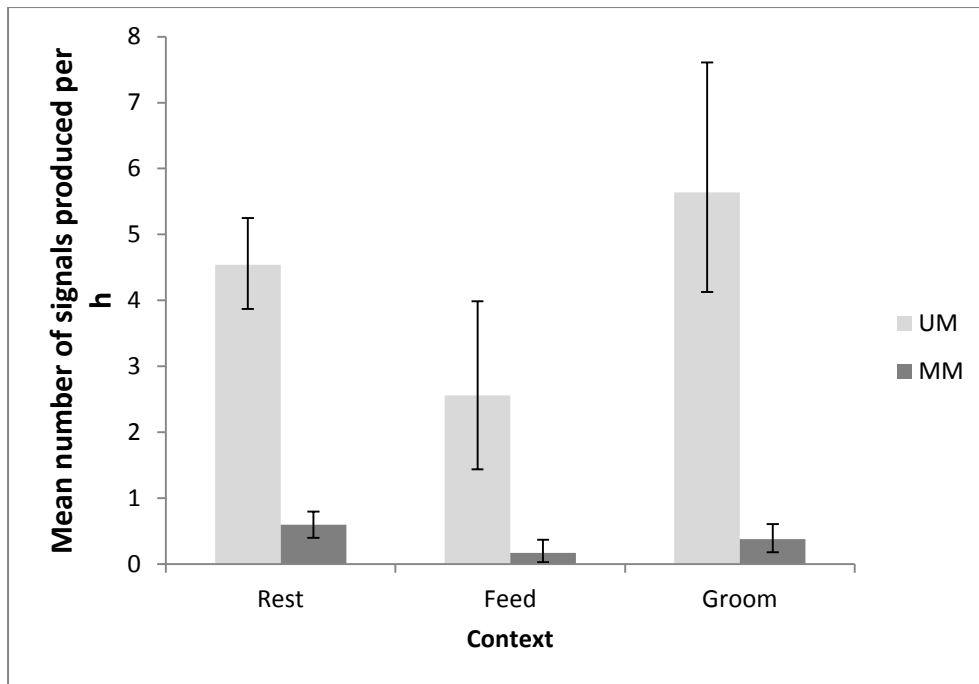
Figure 1. The mean signal production rate (per h) of UM and MM signals in the contexts rest, feed and groom. Error bars represent bootstrapped 95% confidence intervals, based on 1000 iterations. Error bars that do not overlap represent a significant difference. Mean rate and confidence intervals derived 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).

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

	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 (%)
		Across all eight contexts			Across rest, feed, groom and travel contexts					
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

491 The table shows the number of instances and number of individuals observed to produce different MM combinations, both free and fixed, across

492 all eight behavioural contexts, with the range of number of occurrences a combination was produced by a single individual in parentheses. It also

493 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,

494 and of these, the percentage of these occurrences that elicited each of the four recipient response types, or no response. One signal could elicit

495 several responses. Responses were recorded from the start of the signal until 20 s after the end of the signal. The table includes ambiguous

496 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

497 categorized given the signal repertoires used (Table 3). F: facial; V: vocal; G: gestural.

Variation in MM signal production rates: free versus fixed

499   As there is a key cognitive distinction between free MM signal combinations, where signals

500   may be flexibly 'mixed and matched', and fixed MM signal combinations, which are

501   necessarily combined, we investigated the individual and contextual factors that might

502   influence the relative rates of these signals. We constructed a model to test whether variation

503   in the mean MM signal production rate (signals/h) could be explained by interactions

504   between the following fixed factors: type of MM signalling (fixed, free) and (1) context of

505   production, (2) the mean number of individuals in the party, (3) the age of the signaller and

506   (4) the sex of the signaller. The travel context was not included as most MM combinations

507   included facial expressions, which could virtually never be captured during travel.

508

509   The full model comprised these interaction terms and the associated fixed factors. Individual

510   identity was included as a random factor. The dependent variable was mean rate of MM

511   signal production/h. We included 156 data points from 26 individuals in the model. Overall,

512   the full model ($N = 26$ individuals) did not explain a significant amount of variation in MM

513   signal production rates, compared to a null model ($X^2_{11} = 17.06$, $P = 0.106$), indicating that

514   these factors and interactions did not account for significant variation in the MM signal

515   production rates.

516

517   As rank was only available for older individuals, we ran a separate model to assess the effects

518   of rank, by adding rank as a fixed effect and the interaction between rank and signal type to

519   the full model specified above ($N = 114$ data points from 19 individuals). This version of the

520   model that included rank ($N = 19$) confirmed that the overall model did not explain a

521   significant amount of variation in MM signal production rates ($X^2_{13} = 6.70$, $P = 0.917$).

522

523   Rates of the different types of free MM signal combinations were too low and lacked

524   sufficient variability (e.g. the majority of individuals had rates of 0 signals/h) to be subject to

525   inferential statistics; however, Fig. 2 shows that there was interesting variation in the type of

526   MM signals produced in rest, feed and groom contexts.

527
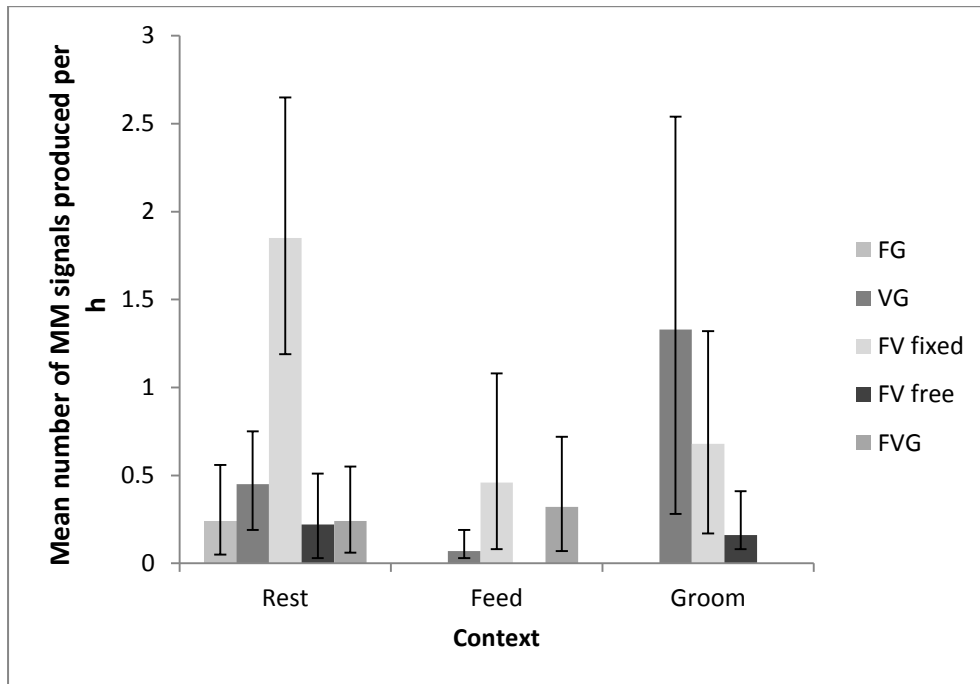
528

529   Figure 2. The MM signal production rate (per h) of facial-gestural (FG), vocal-gestural (VG),

530   facial-vocal fixed (FV fixed), facial-vocal free (FV free) and facial-vocal-gestural (FVG)

531   combinations in the contexts rest, feed and groom. Error bars represent bootstrapped 95%

532   confidence intervals, based on 1000 iterations. Error bars that do not overlap represent a

533   significant difference. MM rate and confidence intervals derived from data of $N = 25$ (facial-

534   gestural, facial-vocal fix, facial-vocal free and facial-vocal-gestural rest), $N = 26$ (vocal-

535   gestural rest, vocal-gestural, facial-vocal fix, facial-vocal-gestural feed and vocal-gestural

536   groom), $N = 23$ (facial-vocal fix and facial-vocal flex groom). Missing bars occur when the

537   MM rate was zero.

538

539

540 *Factors Affecting UM and MM Signal Production*

541 We constructed a model to test whether variation in the mean signal production rate

542 (signals/h) could be explained by interactions between the following fixed factors: type of

543 signalling (UM, MM) and (1) context of production (rest, feed, groom), (2) the mean number

544 of individuals in the party, (3) the age of the signaller and (4) the sex of the signaller. The full

545 model comprised these interaction terms and the associated fixed factors. Individual identity

546 was included as a random factor. The dependent variable was rate of signal production/h. We

547 included 117 data points from 26 individuals in the model. Note that for this model the travel

548 context was excluded as no individuals had sufficient time for UM facial expressions or any

549 MM combination involving facial expressions (i.e. facial-gestural, facial-vocal, facial-vocal-

550 gestural) available in this context.

551

552 Overall, the full model ($N = 26$ individuals) explained a significant amount of variation in

553 signal production rates, compared to a null model ($X^2_{11} = 147.06$, $P < 0.001$). Likelihood

554 ratio tests revealed that there was a significant interaction between type of signal and context

555 ($X^2_2 = 11.12$, $P = 0.004$; Fig. 1). Figure 1 illustrates that signal production rates were

556 significantly higher for UM signals than for MM signals in each context, but the difference

557 between UM and MM rates was greatest in groom and rest contexts, compared to the feed

558 context. No significant interactions between signal type and age ($X^2_1 = 0.26$, $P = 0.613$),

559 signal type and number of individuals in party ($X^2_1 = 2.15$, $P = 0.143$), or signal type and sex

560 ($X^2_1 = 2.47$, $P = 0.116$) were found.

561

562   As rank was only available for older individuals, we ran a separate model to assess the effects

563   of rank, by adding rank as a fixed effect and the interaction between rank and signal type to

564   the full model specified above ($N = 90$ data points from 19 individuals). This version of the

565   model that included rank ($N = 19$) confirmed that the overall model ($X^2_{13} = 138.61$, $P <$

566   0.001) and the Context*UM/MM interaction ($X^2_3 = 19.51$, $P < 0.001$) were significant;

567   however, rank had no significant interaction with signal type ($X^2_1 = 0.08$, $P = 0.784$).

568

569   *Factors Affecting Unimodal Signal Production*

570   As no individuals had sufficient time for UM facial expressions or any MM combination

571   involving facial expressions available to calculate facial expression rates during the travel

572   context, we ran two sets of models to examine (1) the effect of all contexts (including travel)

573   on just vocal and gestural signals (facial expression excluded) and (2) the effect of a reduced

574   set of context (excluding travel) on the full range of signals (facial expression included).

575

576   We first constructed a model to test whether variation in UM signal production rate

577   (signals/h) could be explained by interactions between the following fixed factors: type of

578   UM signal (gestures, vocalizations) and (1) context of production (rest, feed, groom, travel),

579   (2) mean number of individuals in the party, (3) the age of the signaller and (4) the sex of the

580   signaller. The full model comprised the above interaction terms and the associated fixed

581   factors. Individual identity was included as a random factor. The dependent variable was rate

582   of UM signal production/h. This model included the travel context, but excluded facial

583   expressions. We included 184 data points from 26 individuals in the model.

584

585   Overall, the full model ($N = 26$ individuals) explained a significant amount of variation in

586   signal production rates, compared to a null model ($X^2_{13} = 82.24$, $P < 0.001$). Likelihood ratio

587   tests revealed that there was a significant interaction between type of UM signal and context

588   ($X^2_3 = 57.87$, $P < 0.001$; Fig. 3). Figure 3 illustrates that UM gestural signal production rates

589   were significantly higher than UM vocal rates in rest and groom contexts. In contrast, in the

590   travel context, UM vocal signal production rate was significantly higher than gestural

591   production rates. In feed contexts, although vocalizations were given at higher rates than

592   gestures, there was no significant difference between UM modality rates in this context.

593   There were no significant interactions between UM signal type and age ($X^2_1 = 0.04$, $P =$

594   0.843), UM signal type and the mean number of individuals in the party ($X^2_1 = 0.01$, $P =$

595   0.917) or UM signal type and sex ($X^2_1 = 0.92$, $P = 0.338$).

596   As rank was only available for older individuals, we ran a separate model to assess the effects

597   of rank, by adding rank as a fixed effect and the interaction between rank and UM signal type

598   to the full model specified above ($N = 139$ data points from 19 individuals).This version of

599   the model that included rank ($N = 19$) confirmed the overall model ($X^2_{15} = 71.82$, $P < 0.001$),

600   and the Context*Modality of UM signal interaction ($X^2_3 = 46.52$, $P < 0.001$) was significant;

601   however, rank had no significant interaction with modality of the UM signal ($X^2_1 = 0.45$, $P =$

602   0.504).

603
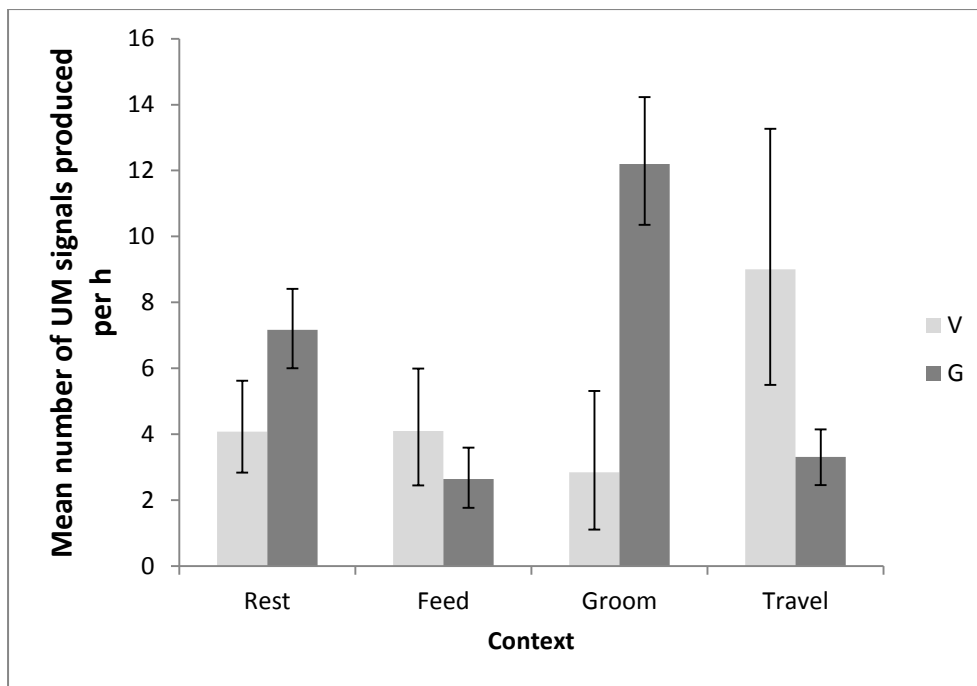
604

605   Figure 3. The UM signal production rate (per h) of vocal (V) and gestural (G) signals in the

606   contexts rest, feed, groom and travel. Error bars represent bootstrapped 95% confidence

607   intervals, based on 1000 iterations. Error bars that do not overlap represent a significant

608   difference. UM rate and confidence intervals derived from data of $N = 25$ (vocalizations and

609   gestures rest), $N = 26$ (vocalizations and gestures feed), $N = 23$ (vocalizations and gestures

610   groom), $N = 20$ (vocalizations travel), $N = 16$ (gestures travel).

611

612   Second, we constructed a model to test whether variation in UM signal production rate

613   (signals/h) could be explained by interactions between the modality/type of UM signal

614   (facial, vocal, gestural) and (1) context of production (rest, feed, groom), (2) mean number of

615   individuals in the party, (3) the age of the signaller and (4) the sex of the signaller. This

616   model excluded the travel context but included facial expressions. The full model comprised

617   the above interaction terms and the associated fixed factors. Individual identity was included

618   as a random factor. The dependent variable was rate of UM signal production/h. We included

619   191 data points from 26 individuals in the model.

620

621   Overall, the full model ($N = 26$ individuals) explained a significant amount of variation in

622   signal production rates, compared to a null model ($X^2_{17} = 144.98$, $P < 0.001$). Likelihood

623   ratio tests revealed that there was a significant interaction between type of UM signal and

624   context ($X^2_4 = 56.84$, $P < 0.001$; Fig. 4). Figure 4 reveals a similar pattern of results as Fig. 3,

625   regarding vocalizations and gestures in rest, feed and groom contexts; however, it also

626   illustrates that the rate of facial signal production was significantly below that for vocal and

627   gestural signals in all three contexts. No significant interactions between UM signal type and

628   age ($X^2_2 < 0.01$, $P = 0.998$), UM signal type and number of individuals in the party ($X^2_2 =$

629   1.05, $P = 0.591$), or UM signal type and sex ($X^2_2 = 2.78$, $P = 0.250$) were found.

630

631   As rank was only available for older individuals, we ran a separate model to assess the effects

632   of rank, by adding rank as a fixed effect and the interaction between rank and UM signal type

633   to the full model specified above ($N = 146$ data points from 19 individuals). This model that

634   included rank ($N = 19$) confirmed the overall model ($X^2_{20} = 129.76$, $P < 0.001$), and the

635   Context*Modality of UM signal interaction ($X^2_4 = 40.26$, $P < 0.001$) remained significant;

636   however, rank had no significant interaction with modality of the UM signal ($X^2_2 = 0.90$, $P =$

637   0.638).

638

639
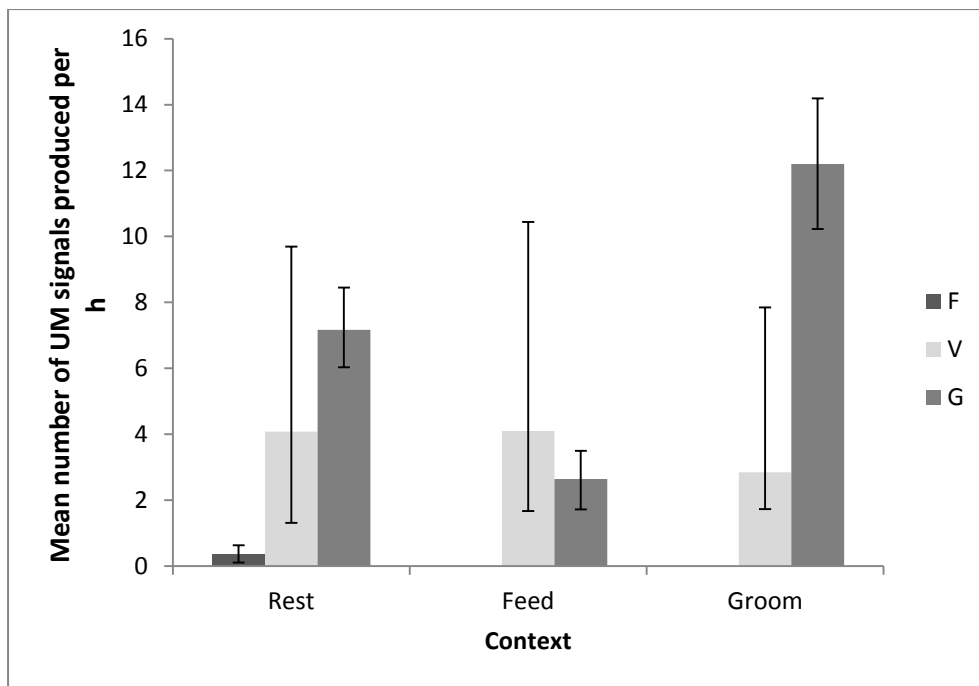
Figure 4. The UM signal production rate (per h) of facial (F), vocal (V) and gestural (G)

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

655   signaller and signal number, which denoted which UM and MM signals were matched

656   together, were included as random factors. There were 104 data points from seven individuals

657   in the model.

658

659   Overall, the full model ($N = 7$ individuals) explained a significant amount of variation in

660   whether or not the focal individual's signal elicited a response from recipients within 5 m,

661   compared to a null model ($X^2_4 = 37.12$, $P < 0.001$). Likelihood ratio tests revealed that there

662   was a significant main effect of signal type produced ($X^2_2 = 34.16$, $P < 0.001$; Fig. 5). Figure

663   5 shows that UM vocal signals were significantly less likely to elicit a response from

664   recipients than MM vocal-gestural signals or UM gestural signals. There was no significant

665   difference in the proportion of MM vocal-gestural and UM gestural signals that elicited a

666   response. A trend for lower ranking individuals to be more likely to receive a response than

667   higher ranking individuals was found, but this effect was not significant ($X^2_1 = 2.85$, $P =$

668   0.092), nor was the effect of the number of individuals within 5 m of the focal individual (X^2_1

669    $= 2.61$, $P = 0.106$).

670
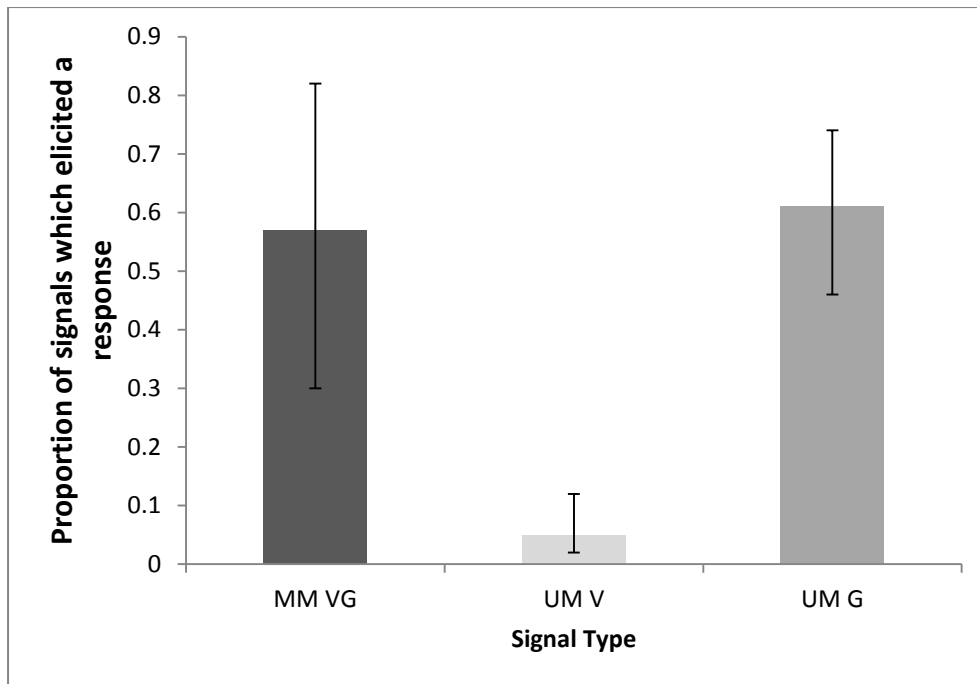
671

672   Figure 5. The mean proportion of focal individual MM vocal-gestural (VG), UM vocal (V)

673   and UM gestural (G) signals that elicited a response from recipients within 5 m. Error bars

674   represent bootstrapped 95% confidence intervals, based on 1000 iterations. Error bars that do

675   not overlap represent a significant difference. Data based on $N = 7$ individuals.

676

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

678   We investigated whether the main responses elicited by a MM vocal-gestural signal matched

679   those elicited by either of its UM components. Main responses were defined as the most

680   active response that was the closest to the final behavioural outcome. For instance, if in

681   response to a focal individual signal, a recipient looked at the signaller, approached and

682   groomed, the main response was taken to be grooming.

683   Of the seven individuals for whom we compared MM signals and their UM components,

684   Table 6 shows the four individuals from whom the MM signal elicited a response, and thus

685   the responses to the UM components could be compared to the response to the MM signal

686 (see Supplementary Table S4 for main responses elicited from all MM and matched UM

687 signals, including those that did not elicit a response/were ignored). While MM signals from

688 two female individuals elicited responses that matched the majority of responses to their UM

689 gestural but not UM vocal signals (dominance of gestural response), one male individual

690 elicited different responses to his MM signal than either of the components (emergence).

691

692 Table 6. Instances where the main response of the UM vocal signal and UM gestural signal

693 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)

694 See Supplementary Table S4 for details of the type of responses elicited. Only individuals

695 whose MM signals received a response were included in this table.

696

697 **DISCUSSION**

698 Although MM signals may not be as common as UM signals, this study has documented the

699 production of 48 different free MM combinations. While 22 of 26 individuals produced at

700   least one free MM signal, each broad type of MM signal combination was observed to be

701   produced by at least nine of the 26 focal individuals. This suggests that the vast majority of

702   individuals have the capacity and motivation to flexibly and simultaneously combine signals

703   from different modalities, albeit rarely. In the future, a largescale, collaborative approach to

704   document free MM signal production across individuals' life spans and across study sites

705   would shed valuable light on the mechanisms underpinning the production of these

706   combinations (e.g. social learning, individual innovation, innately predetermined) and

707   whether cultural variation exists in the type of free MM signal combinations commonly

708   produced.

709   The number of UM signals produced per h was found to be more than 10 times higher than

710   the number of MM signals in our study. This contrasts sharply with the two previous captive

711   studies to compare proportions of UM and MM signals, which both found much higher

712   relative rates of MM signals. Pollick and de Waal's (2007) chimpanzee signals consisted of,

713   56% gestures, 22.5% facial/vocal signals and 21.6% MM combinations of the two. Similarly,

714   Taglialatela et al. (2015) found that approximately half of the vocalizations recorded were

715   accompanied by another communicative signal/behaviour. The relatively low levels of MM

716   signals to UM signals, compared to these previous studies (see Supplementary Fig. S5 and

717   Table S6), could be attributable to several factors. First, we identified MM signals as ones

718   with temporal overlap, whereas previous studies considered signals or behaviours produced

719   within 2 s (Taglialatela et al., 2015) or 10 s (Pollick & de Waal, 2007) of each other as MM

720   signals. Second, we only considered vocal, gestural and facial signals, whereas Taglialatela et

721   al. (2015) also included combinations of behaviours such as play or chase with vocalizations

722   to be MM signals. Third, our study shows the importance of context in influencing the

723   relative rate of UM and MM signals, whereas previous studies did not examine the same

724   contexts as this study, nor did they specifically compare different contexts. For instance,

725   Pollick & de Waal did not include rest, but importantly did include social play. Play is highly

726   interactive, and it is common for individuals to show a range of MM signals in this context,

727   such as play face, laughter, and various manual and nonmanual gestures simultaneously, so

728   this could also explain the higher proportion of MM signals recorded. Finally, these previous

729   two studies were conducted in captivity, where the social and physical environment may

730   favour higher rates of MM signalling. In an enclosed area individuals are not normally able to

731   express fission –fusion behaviour and this may mean that individuals need to use more

732   sophisticated signals to negotiate tense social interactions, where in the wild they could

733   simply leave the party, or seek a greater distance from certain individuals. Furthermore, in a

734   captive environment visibility is generally much higher than in the dense tropical rainforest,

735   meaning that MM combinations including visual signals are more likely to be successfully

736   received. Investigating MM communication in wild savannah chimpanzees could be an

737   interesting avenue for future research to explore whether the differences between the levels of

738   MM signals produced in the wild and captivity seen so far is due to strategies learnt to cope

739   with the limited space in captivity and interactions with humans, or in fact the level of

740   visibility.

741   Our results partially supported our hypotheses that MM signals would be more likely to elicit

742   a response than UM signals: the likelihood of a response was significantly higher with a MM

743   grunt + gesture signal than a UM grunt signal, but similar to the matching UM gesture signal.

744   This suggests that in the context of these specific signals, adding a vocal signal to a gesture

745   does not change the likelihood of eliciting a response; in contrast, adding a gesture signal to a

746   vocalization significantly improves the chances of eliciting a response. This supports findings

747   from Pollick & de Waal's (2007) study that indicated that MM signals of gestures combined

748   with a vocalization or facial expression were no more effective at eliciting responses than

749   gestures alone. Although adding vocalizations to gestures may not increase the likelihood of

750 obtaining a response, it may help disambiguate the signaller's intended meaning or convey

751 more information than the UM signals in isolation. Indeed, in one individual the responses

752 elicited to the MM signal were different to both the vocal and gestural components,

753 indicating MM signals in chimpanzees have the potential to have emergent functions.

754 Equally, it could be the case here that vocalizations are used as attention-getting signals

755 alongside gestures (similar to Leavens, Russell & Hopkins, 2010), for example when the

756 recipient does not have their visual attention directed towards the signaller. In this case the

757 gesture might be the signal to which the signaller actually wants the recipient to attend.

758 Descriptively, when examining MM signals that elicited a recipient response, in two of the

759 four individuals the type of response elicited by the MM signal was more likely to match the

760 response elicited by the gestural than the vocal components. Viewed in the framework of

761 Partan and Marler (2005), this indicates that for these individuals this particular free MM

762 signal may be best characterized as a nonredundant combination that retains a dominant

763 gestural response. Whether similar findings would be obtained if a different type of

764 vocalization had been focused on remains unclear. The grunt vocalization investigated here

765 might be relatively ambiguous as it is frequently produced in a variety of contexts. In

766 contrast, many of the gestures that were highly successful in eliciting responses (e.g. present

767 groom) were highly specific to a groom context and had a clear and measurable recipient

768 response. As our repertoire illustrates, wild chimpanzees produce a large array of free MM

769 signals and further research needs to systematically investigate the recipient responses to

770 these and their matched UM component signals in order to understand the range of functions

771 free MM signals have in this species.

772 In contrast to our predictions, we found that the proportion of signals that elicited a response

773 was not dependent on the rank of the individual who produced the signal, nor the number of

774 individuals who were within 5 m of this individual. It could be that the likelihood of a

775 response may be more influenced by the rank difference or degree of friendship between

776 signaller and recipient rather than the absolute rank of the signaller. We were not able to

777 accurately calculate such relative dyadic measures, as for the majority of signals it was

778 difficult to discern which individual was the recipient, and potentially there could have been

779 several. In terms of the number of potential receivers, it could be the case that the majority of

780 signals are in fact directed at a specific individual (e.g. Schel, Machanda et al., 2013), even if

781 this might be difficult for human observers to detect, and thus the number of other individuals

782 in the vicinity may not be an important predictor of a response. In the grunt + gesture MM

783 signals that we investigated this is likely to be particularly true, as most signals occurred in a

784 groom context, where the signals are likely to be directed at the grooming partner. In

785 addition, we only considered grunts, which are an example of a 'proximal' vocalization that

786 Taglialatela et al (2009) showed were more likely to be directed towards specific individuals,

787 and to be processed differently by recipients, compared to 'broadcast' vocalizations, such as

788 pant hoots.

789 Relative rates of vocal, gestural and facial signal production varied as a function of context. It

790 was predicted that wild chimpanzees might tailor their signalling to the recipient's attentional

791 state, as has been shown in captivity (Leavens et al., 2010), and there were indications of this

792 in this wild population. The signal production rate of UM gestures was found to be

793 significantly higher than UM vocalizations in rest and groom contexts. This could be because

794 in these contexts the focal individual might be more likely to have the visual attention of the

795 recipient (especially when in close proximity, such as during grooming), whereas during feed

796 or travel recipients are less likely to have the visual attention of others. Conversely, during

797 travel individuals appear to produce significantly higher rates of vocal signals, which are

798 more likely to be received not only by members of their own party, but also by more distantly

799 located individuals. Facial expressions were only observed in a rest context; for feed and

800 groom contexts the 11 and nine individuals that met the time criteria for calculation of a

801 signalling rate had a facial signal production rate of zero. Thus UM facial expressions were

802 recorded very rarely, and rates were significantly below those of vocal and gestural signals in

803 rest, feed and groom contexts. Also note that we only coded salient facial expressions (see

804 Table 3) and had we applied full FACS coding (Vick, Waller, Parr, Pasqualini & Bard, 2007)

805 to our videos, subtler facial movements might have been captured. Nevertheless, in the dense

806 forest environment facial expressions alone may be difficult for receivers to detect, and they

807 may be more effective when combined with other signals. Indeed, Fig. 2 shows that facial

808 expressions are produced in all contexts in which we could measure them in combination

809 with other signals. This highlights the importance of adopting a holistic MM approach to

810 studying animal communication: facial expressions in wild chimpanzees are most commonly

811 produced as part of MM signals and extracting facial expressions and analysing them in

812 isolation from the composite signal is likely to lead to incorrect understanding of signal

813 function.

814 Contrary to our predictions, none of the individual factors we included in our models

815 influenced the rate of UM and MM signal production. Age, sex and rank did not significantly

816 interact with UM/MM signal rate. This indicates that learning to combine signals from

817 different modalities and how to effectively use different types of signals may occur relatively

818 early in development, before early adulthood. Previous research has shown that infant and

819 juvenile chimpanzees may produce several different gestural signals in sequences as a 'fail-

820 safe' strategy to elicit a response. In contrast, more mature individuals were found to produce

821 fewer, but more successful signals (Hobaiter & Byrne, 2011b), and no differences in gesture

822 signalling strategies were seen between subadults (10–14/15 years) and adults. In line with

823 these findings, we found that age of the sub-adult and adult individuals we studied did not

824 influence the relative proportion of MM and UM signals produced. As Hobaiter & Byrne

825 (2011b) found juveniles and infants often used rapid fire gesture sequences, probably to

826 encourage recipients to respond, future research should investigate whether infants and

827 juveniles adopt a similar strategy with signal combinations and produce a higher proportion

828 of MM signals than adults.

829 By examining multiple modalities and their combinations simultaneously we have revealed

830 free MM combinations and flexible usage of different types of UM signals across contexts.

831 Facial expressions were rarely produced in isolation and instead were more commonly

832 combined with other signals: artificially extracting facial expressions from these composite

833 signals could lead to misunderstanding of signal function. We advocate a MM approach to

834 gain a full understanding not only of animal communication, but also of the evolutionary

835 roots of human language. Human language is a multimodal communication system, with

836 gestures and facial expressions accompanying and modifying the meaning of speech, and this

837 study has shown that the ability and motivation to flexibly combine different signals are

838 present in wild chimpanzees, and thus are likely to be present in our last common ancestor.

839 Further research into the function of different free MM combinations may reveal the potential

840 for generativity (emergent function, Partan & Marler, 2005) and social learning of MM signal

841 combinations, which would have significant impact on our understanding of the evolution of

842 these key facets of language.

843 In conclusion, our results reveal an impressive repertoire of free MM signals, but that these

844 signals are used rarely compared to gestures and vocalizations in isolation. Interestingly,

845 facial expressions are more commonly produced as part of MM signals than in isolation in

846 several contexts. Systematic investigation of the MM grunt + gesture signal and the UM

847 component signals revealed MM signals were more likely to elicit a response than UM vocal

848 signals, but not UM gestural signals, and several potential functions for this specific type of

849 MM signal were identified. The relative rates of UM vocal, gestural and facial signals varied

850   across contexts, indicating flexible use of different signalling modalities across contexts. The

851   flexibility in communicative signalling this study has revealed, by adopting a MM approach,

852   may represent an important cognitive foundation from which our own complex multimodal

853   communication system could have evolved.

854

866

867   **Supplementary Material**

868

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

870

871   **REFERENCES**

872   Altmann, J. (1974). Oberservational study of behaviour: sampling methods.

873   *Behaviour, 49(3)*, 227-267. http://dx.doi.org/10.1163/156853974X00534.

874   Bates, D., & Maechler, M. (2009). Dai B lme4: Linear mixed-effects models using

875   S4 classes. R package version 0.999375-28. http://cran.R-project.org/package=lme4.

876   Chapman, C. A., & Wrangham, R. W. (1993). Range use of the forest chimpanzees

877   of Kibale: implications for the understanding of chimpanzee social

878   organization. *American Journal of Primatology*, *31*(4), 263-273.

879   Cooper, B. G., & Goller, F. (2004). Multimodal signals: enhancement and

880   constraint of song motor patterns by visual display. *Science*, *303*(5657), 544-546.

881   http://dx.doi.org/10.1126/science.1091099.

882   Crawley, M. J. (2002). *Statistical computing: an introduction to data analysis*

883   *using S-Plus.* Chichester, U.K.: Wiley.

884   Crockford, C., Wittig, R. M., & Zuberbühler, K. (2015). An intentional vocalization

885   draws others' attention: A playback experiment with wild chimpanzees. *Animal*

886   *Cognition*, *18*(3), 581-591. http://dx.doi.org/10.1007/s10071-014-0827-z.

887   de Luna, A. G., Hödl, W., & Amezquita, A. (2010). Colour, size and movement as

888   visual subcomponents in multimodal communication by the frog Allobates

889   femoralis. *Animal Behaviour*, *79*(3), 739-745.

890   http://dx.doi.org/10.1016/j.anbehav.2009.12.031.

891   de Vries, H. A. N., Stevens, J. M., & Vervaecke, H. (2006). Measuring and testing

892   the steepness of dominance hierarchies. *Animal Behaviour*, *71*(3), 585-592.

893   http://dx.doi.org/10.1016/j.anbehav.2005.05.015.

894   Faraway, J. J. (2006). Extending the Linear Model with R: Generalized

895   Linear. *Mixed Effects and Nonparametric Regression Models*. Boca Raton, FL: CRC

896   Press.

897   Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability

898   studies. *Applied Psychological Measurement*, *5*(1), 105-112.

899   Georgiev, A. V., Russell, A. F., Thompson, M. E., Otali, E., Muller, M. N., &

900   Wrangham, R. W. (2014). The foraging costs of mating effort in male chimpanzees (*Pan

901   troglodytes schweinfurthii*). *International Journal of Primatology*, *35*(3-4), 725-745.

902   http://dx.doi.org/10.1007/s10764-014-9788-y.

903   Gruber, T., & Zuberbühler, K. (2013). Vocal recruitment for joint travel in wild

904   chimpanzees. *PLoS One*, *8*(9), e76073. http://dx.doi.org/10.1371/journal.pone.0076073.

905   Hayashi, M. (2007). A new notation system of object manipulation in the nesting-

906   cup task for chimpanzees and humans. *Cortex*, *43*(3), 308-318.

907   http://dx.doi:10.1016/S0010-9452(08)70457-X.

908   Hebets, E. A. (2008). Seismic signal dominance in the multimodal courtship display

909   of the wolf spider Schizocosa stridulans Stratton 1991. *Behavioral Ecology*, *19*(6), 1250-

910   1257. http://dx.doi.org/10.1093/beheco/arn080.

911   Hebets, E. A., & Papaj, D. R. (2005). Complex signal function: developing a

912   framework of testable hypotheses. *Behavioral Ecology and Sociobiology*, *57*(3), 197-214.

913   http://dx.doi.org/10.1007/s00265-004-0865-7.

914   Hebets, E. A., & Uetz, G. W. (1999). Female responses to isolated signals from

915   multimodal male courtship displays in the wolf spider genus *Schizocosa* (Araneae:

916   Lycosidae). *Animal Behaviour*, *57*(4), 865-872.

917   Higham, J. P., & Hebets, E. A. (2013). An introduction to multimodal

918   communication. *Behavioral Ecology and Sociobiology*, *67*(9), 1381-1388.

919   http://dx.doi.org/10.1007/s00265-013-1590-x.

920   Hobaiter, C., & Byrne, R. W. (2011a). The gestural repertoire of the wild

921   chimpanzee. *Animal Cognition*, *14*(5), 745-767. http://dx.doi.org/10.1007/s10071-011-

922   0409-2.

923   Hobaiter, C., & Byrne, R. W. (2011b). Serial gesturing by wild chimpanzees: its

924   nature and function for communication. *Animal Cognition*, *14*(6), 827-838.

925   http://dx.doi.org/10.1007/s10071-011-0416-3.

926   Hopkins, W. D., Taglialatela, J. P., & Leavens, D. A. (2011). Do chimpanzees have

927   voluntary control of their facial expressions and vocalizations. In A. Villain, J-L.

928   Schwartz, C. Abry & J. Vauclair (Eds), *Primate communication and human language:*

929   *Vocalisation, gestures, imitation and deixis in humans and non-humans*, vol. *1*, pp. 71-88.

930   Amsterdam, Netherlands: John Benjamins. http://dx.doi.org/10.1075/ais.1.05hop.

931   Kalan, A. K., & Boesch, C. (2015). Audience effects in chimpanzee food calls and

932   their potential for recruiting others. *Behavioral Ecology and Sociobiology*, *69*(10), 1701-

933   1712. http://dx.doi.org/10.1007/s00265-015-1982-1.

934   Leavens, D. A., Hopkins, W. D., & Bard, K. A. (1996). Indexical and referential

935   pointing in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, *110*(4),

936   346.

937   Leavens, D. A., Hopkins, W. D., & Thomas, R. K. (2004). Referential

938   communication by chimpanzees (*Pan troglodytes*). *Journal of Comparative*

939   *Psychology*, *118*(1), 48. http://dx.doi.org/10.1037/0735-7036.118.1.48.

940   Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2010). Multimodal

941   communication by captive chimpanzees (*Pan troglodytes*). *Animal Cognition*, *13*(1), 33-

942   40. http://dx.doi.org/10.1007/s10071-009-0242-z.

943   Liebal, K., Waller, B. M., Slocombe, K. E., & Burrows, A. M. (2013). *Primate*

944   *Communication: a multimodal approach*. Cambridge, U.K.: Cambridge University Press.

945   http://dx.doi.org/10.1017/CBO9781139018111 .

946   Muller, M. N., & Wrangham, R. W. (2014). Mortality rates among Kanyawara

947   chimpanzees. *Journal of Human Evolution*, *66*, 107-114.

948   http://dx.doi.org/10.1016/j.jhevol.2013.10.004

949   Nishida, T., Kano, T., Goodall, J., McGrew, W. C., & Nakamura, M. (1999).

950   Ethogram and Ethnography of Mahale Chimpanzees. *Anthropological Science*, *107*(2),

951   141-188. http://dx.doi.org/10.1537/ase.107.141.

952   Parr, L. A. (2004). Perceptual biases for multimodal cues in chimpanzee (*Pan*

953   *troglodytes)* affect recognition. *Animal Cognition*, *7*(3), 171-178.

954   http://dx.doi.org/10.1007/s10071-004-0207-1.

955   Parr, L. A., Waller, B. M., Vick, S. J., & Bard, K. A. (2007). Classifying

956   chimpanzee facial expressions using muscle action. *Emotion*, *7*(1), 172.

957   http://dx.doi.org/10.1037/1528-3542.7.1.172.

958   Partan, S. R. (2002). Single and multichannel signal composition: facial expressions

959   and vocalizations of rhesus macaques (*Macaca mulatta*). *Behaviour*, *139*(8), 993-1027.

960   http://dx.doi.org/10.1163/15685390260337877.

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

962   1272-1273.

963   Partan, S. R., & Marler, P. (2005). Issues in the classification of multimodal

964   communication signals. *The American Naturalist*, *166*(2), 231-245.

965   http://dx.doi.org/10.1086/431246.

966   Partan, S. R., Fulmer, A. G., Gounard, M. A., & Redmond, J. E. (2010).

967   Multimodal alarm behavior in urban and rural gray squirrels studied by means of

968   observation and a mechanical robot. *Current Zoology*, *56*(3), 313-326.

969   Partan, S. R., Larco, C. P., & Owens, M. J. (2009). Wild tree squirrels respond with

970   multisensory enhancement to conspecific robot alarm behaviour. *Animal*

971   *Behaviour*, *77*(5), 1127-1135. http://dx.doi.org/10.1016/j.anbehav.2008.12.029.

972   Pollick, A. S., & De Waal, F. B. (2007). Ape gestures and language

973   evolution. *Proceedings of the National Academy of Sciences*, *104*(19), 8184-8189.

974   http://dx.doi.org/10.1073/pnas.0702624104.

975   Rowe, C. (1999). Receiver psychology and the evolution of multicomponent

976   signals. *Animal Behaviour*, *58*(5), 921-931. http://dx.doi.org/10.1006/anbe.1999.1242.

977   Schel, A. M., Machanda, Z., Townsend, S. W., Zuberbühler, K., & Slocombe, K. E.

978   (2013). Chimpanzee food calls are directed at specific individuals. *Animal*

979   *Behaviour*, *86*(5), 955-965. http://dx.doi.org/10.1016/j.anbehav.2013.08.013.

980   Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., & Slocombe, K. E.

981 (2013). Chimpanzee alarm call production meets key criteria for intentionality. *PLoS*

982 *One*, *8*(10), e76674. http://dx.doi.org/10.1371/journal.pone.0076674.

983   Schwarz, A. (1974). Sound production and associated behaviour in a cichlid fish,

984 *Cichlasoma centrarchus*. *Zeitschrift fuer Tierpsychologie*, *35*(2), 147-156.

985   Slocombe, K. E., Kaller, T., Turman, L., Townsend, S. W., Papworth, S., Squibbs,

986 P., & Zuberbühler, K. (2010). Production of food-associated calls in wild male

987 chimpanzees is dependent on the composition of the audience. *Behavioral Ecology and*

988 *Sociobiology*, *64*(12), 1959-1966. http://dx.doi.org/10.1007/s00265-010-1006-0.

989   Slocombe, K. E., Waller, B. M., & Liebal, K. (2011). The language void: the need

990 for multimodality in primate communication research. *Animal Behaviour*, *81*(5), 919-924.

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

992   Slocombe, K. E., & Zuberbühler, K. (2005). Functionally referential

993 communication in a chimpanzee. *Current Biology*, *15*(19), 1779-1784.

994 http://dx.doi.org/10.1016/j.cub.2005.08.068.

995   Slocombe, K. E., & Zuberbühler, K. (2006). Food-associated calls in chimpanzees:

996 responses to food types or food preferences? *Animal Behaviour*, *72*(5), 989-999.

997 http://dx.doi.org/10.1016/j.anbehav.2006.01.030.

998   Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment

999 screams as a function of audience composition. *Proceedings of the National Academy of*

1000 *Sciences*, *104*(43), 17228-17233. http://dx.doi.org/10.1073/pnas.0706741104.

1001   Slocombe, K. E., & Zuberbühler, K. (2010). Vocal communication in

1002 chimpanzees. In E. V. Lonsdorf, S. R. Ross & T. Matsuzawa (Eds.) *The mind of the*

1003   *chimpanzee: ecological and experimental perspectives* (pp. 192-207). Chicago, IL:

1004   University of Chicago Press.

1005   Smith, W. J. (1977). The behavior of communicating: an ethological approach.

1006   Harvard University Press, Cambridge, MA.

1007   Smith, C. L., & Evans, C. S. (2013). A new heuristic for capturing the complexity

1008   of multimodal signals. *Behavioral Ecology and Sociobiology*, *67*(9), 1389-1398.

1009   http://dx.doi.org/10.1007/s00265-013-1490-0.

1010   Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2009).

1011   Visualizing vocal perception in the chimpanzee brain. *Cerebral Cortex*, *19*(5), 1151-

1012   1157. http://dx.doi.org/10.1093/cercor/bhn157.

1013   Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2011).

1014   Chimpanzee vocal signaling points to a multimodal origin of human language. *PloS*

1015   *one*, *6*(4), e18852. http://dx.doi.org/10.1371/journal.pone.0018852.

1016   Taglialatela, J. P., Russell, J. L., Pope, S. M., Morton, T., Bogart, S., Reamer, L. A.,

1017   Schapiro, S. J., & Hopkins, W. D. (2015). Multimodal communication in

1018   chimpanzees. *American journal of primatology*, *77*(11), 1143-1148.

1019   http://dx.doi.org/10.1002/ajp.22449.

1020   Tomasello, M. (2008). *The origins of human communication*. Cambridge, MA: MIT

1021   Press.

1022   Townsend, S. W., & Züberbuhler, K. (2009). Audience effects in chimpanzee

1023   copulation calls. *Communicative & integrative biology*, *2*(3), 282-284.

1024   http://dx.doi.org/10.4161/cib.2.3.6796.

1025   Uetz, G. W., & Roberts, J. A. (2002). Multisensory cues and multimodal

1026   communication in spiders: insights from video/audio playback studies. *Brain, Behavior*

1027   *and Evolution*, *59*(4), 222-230. http://dx.doi.org/10.1159/000064909.

1028   Uetz, G. W., Roberts, J. A., & Taylor, P. W. (2009). Multimodal communication

1029   and mate choice in wolf spiders: female response to multimodal versus unimodal

1030   signals. *Animal Behaviour*, *78*(2), 299-305.

1031   http://dx.doi.org/10.1016/j.anbehav.2009.04.023.

1032   Waller, B. M., Liebal, K., Burrows, A. M., & Slocombe, K. E. (2013). How can a

1033   multimodal approach to primate communication help us understand the evolution of

1034   communication? *Evolutionary Psychology*, *11*(3), 147470491301100305.

1035   http://dx.doi.org/10.1177/147470491301100305.

1036   Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng,

1037   L., ... & Slocombe, K. E. (2015). Vocal learning in the functionally referential food grunts

1038   of chimpanzees. *Current Biology*, *25*(4), 495-499.

1039   http://dx.doi:10.1016/j.cub.2014.12.032.

1040   Vick, S. J., Waller, B. M., Parr, L. A., Pasqualini, M. C. S., & Bard, K. A. (2007).

1041   A cross-species comparison of facial morphology and movement in humans and

1042   chimpanzees using the facial action coding system (FACS). *Journal of Nonverbal*

1043   *Behavior*, *31*(1), 1-20. http://dx.doi.org/10.1007/s10919-006-0017-z.

1044   Wilson, M. L., Kahlenberg, S. M., Wells, M., & Wrangham, R. W. (2012).

1045   Ecological and social factors affect the occurrence and outcomes of intergroup encounters

1046   in chimpanzees. *Animal Behaviour*, *83*(1), 277-291.

1047   http://dx.doi:10.1016/j.anbehav.2011.11.004.

1048   Wrangham, R. W., Clark, A. P., & Isabirye-Basuta, G. (1992). Female social

1049   relationships and social organization of Kibale Forest chimpanzees. *Topics in*

1050   *primatology*, *1*, 81-98.