

# Use and Function of Multi-Articulator Versus Multi-Sensory Acts in the Close-Range Communication of Orang-Utans

Marlen Fröhlich (✉ [marlen.froehlich@uzh.ch](mailto:marlen.froehlich@uzh.ch))

University of Zurich <https://orcid.org/0000-0002-1948-7002>

**Natasha Bartolotta**

University of Zurich

**Caroline Fryns**

University of Zurich

**Colin Wagner**

University of Strasbourg

**Laurene Momon**

University of Strasbourg

**Marvin Jaffrezic**

University of Strasbourg

**Tatang Mitra Setia**

Universitas Nasional

**Maria van Noordwijk**

University of Zurich

**Carel van Schaik**

University of Zurich

---

## Article

**Keywords:** Multiplex communication, multimodal communication, behavioural plasticity, socio-ecological environment, orang-utan, *Pongo abelii*, *Pongo pygmaeus*

**Posted Date:** December 16th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-123298/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Communications Biology on July 27th, 2021. See the published version at <https://doi.org/10.1038/s42003-021-02429-y>.

# Abstract

From early infancy, human face-to-face communication is “multimodal”, comprising a plethora of interlinked articulators and sensory modalities. Although there is also growing evidence for this in nonhuman primates, the functions of integrating articulators (i.e. multiplex or multi-articulator acts) and channels (i.e. multimodal or multi-sensory acts) remain poorly understood. Here, we studied close-range social interactions within and beyond mother-infant pairs of Bornean and Sumatran orang-utans living in wild and captive settings, to examine to what extent species, setting and recipient-dependent factors affected the use of and responses to multi-sensory as well as multi-articulator communication. Results showed that both multi-sensory and multi-articulatory acts were more effective at eliciting responses (i.e. “apparently satisfactory outcomes”) than their respective uni-component parts, and generally played a larger role in wild populations. However, only multi-articulator acts were used more when the presumed goal did not match the dominant outcome for a specific communicative act, and were more common among non-mother-infant dyads and Sumatrans across settings. We suggest that communication through multiple sensory channels primarily facilitates effectiveness, whereas a flexible combination of articulators is relevant when social tolerance and interaction outcomes are less predictable. These different functions underscore the importance of distinguishing between these forms of multi-component communication.

## Introduction

Human face-to-face communication is a multi-component phenomenon: our everyday speech is embedded in an interactional exchange of unified visual, auditory, and often even tactile signals. Some parts of these multimodal displays are intrinsically coupled due to the effort of vocal production (such as mouth movement accompanying speech sounds), but others are flexible (e.g. gaze and co-speech gestures). Research on the nature and function of human multi-component interaction has recently focused particularly on flexible combinations of different articulators (i.e. communication organs such as hands, lips and eyes) e.g. <sup>1,2</sup>. For instance, speech acts accompanied by gestures and gaze are processed faster <sup>3,4</sup> and elicit faster responses <sup>5</sup>, respectively. This suggests that a complex orchestration of articulators and sensory channels facilitates comprehension and prediction during language processing <sup>6</sup>.

Many non-human species also have a natural predisposition for multi-component social interactions, as evident in complex mating, warning and dominance displays <sup>7,8</sup>. Multi-component signalling can be distinguished based on the perspective of production versus perception, as emphasized recently by Holler and Levinson <sup>6</sup>: “multiplex” communication involves at least two different articulators or communication organs at the production side <sup>6</sup>, such as hands plus gaze, whereas multimodal communication involves at least two different sensory channels at the perception end, such as visual plus auditory <sup>9</sup>. Many multi-component acts are both multiplex and multimodal, for instance a tactile gesture combined with a facial expression, whereas some are just multimodal, such as the audio-visual *loud scratch* gesture <sup>10,11</sup>, and

others are only multiplex, such as a visual gesture combined with a facial expression. In fact, our closest living relatives, the great apes, are renowned for signalling intentional communicative acts in large part by non-vocal means in their close range dyadic interactions<sup>8,12,13</sup>. Not only are many of these signals intrinsically multimodal (e.g. tactile gestures that can be simultaneously seen and felt by a receiver, or lip-smacking which can be seen and heard), but they can also be integrated with other, non-vocal or vocal means in multimodal signal combinations e.g.<sup>12,13,14</sup>. Because the term “multimodality” has confusingly been used for both types of multi-component communication<sup>15</sup>, we will henceforth refer to multi-sensory and multi-articulator acts for multimodal and multiplex (in the sense of ref. 7), respectively.

The fact that close-range communicative acts may be either multi-sensory or multi-articulatory (even if many are both) highlights the importance of assessing whether they serve different communicative functions. However, to date no study has explicitly investigated and compared the usage of uni-/multi-sensory versus uni-/multi-articulator communicative acts in a great ape taxon (nor, to our knowledge, in humans). The theoretical and empirical differences between these combination types are often ignored in comparative research<sup>12,15</sup>, but addressing them would be key to draw conclusions about homologous features in the human/ape communication system<sup>16</sup>.

A neurobiological perspective underscores the plausibility of this differentiation: in multi-sensory communication, the recipient is forced to integrate incoming information in at least two different sensory channels that initially are processed in different brain regions. Visual and auditory pathways, for instance, are largely separate before converging in the ventrolateral prefrontal cortex (vlPFC) onto neurons that represent higher-order multisensory representations of signals, such as vocalizations and their associated facial expressions<sup>17</sup>. This need to integrate may make it more likely that the communicative act is accurately processed, suggesting that multi-sensory communication serves to ensure that a signal is understood<sup>18,19</sup>.

The multi-articulator case explicitly takes the signaller’s perspective. In contexts or situations requiring a multi-articulator act, the signaller is forced to execute (at least) two different motor commands in different articulators. For instance, neurobiological research on human multimodal processing suggests that the integration between speech and gesture depends on context and is under voluntary control rather than obligatory<sup>20</sup>: co-speech gestures may provide additional information depending on the communicative nature of the situation (e.g. whether or not there is shared common ground between the signaller and the recipient)<sup>21</sup>, as well as on gaze direction (i.e. whether or not the signaller’s gaze is directed at the addressee)<sup>22</sup>. Together with rich evidence that multi-articulator acts serve to refine messages<sup>1,23,24</sup>, this suggests they are of particular relevance when outcomes (due to lower degrees of familiarity and social tolerance) are less predictable.

These neurobiological considerations suggest that multi-sensory and multi-articulator acts may serve different functions. Comparative researchers have recently begun to study the function of great apes’ multi-component communication via observational research, focusing on bi-articulatory gesture-vocal

combinations<sup>12-14,25</sup> and considering two different major function(s): redundancy and refinement<sup>9,15,19</sup>. The redundant signal (hereafter referred to as 'redundancy') hypothesis implies that the different components convey the same information<sup>9</sup>, facilitating the detection and processing of a message<sup>18</sup>. In contrast, the refinement hypothesis posits that the presence of one signal component may provide the context in which a receiver can interpret and respond to the second, with the combinations serving to disambiguate meanings (i.e. functions) when these partly overlap<sup>15,19</sup>. An important shortcoming of previous work, however, was that researchers did not tease apart production and perception of communicative acts, or whether constituent parts varied with regard to articulators (body parts) or sensory channels (modalities). Teasing these apart will allow us to gain more insight into the function of multi-component communication in great apes.

The aim of this study was to disentangle multi-sensory and multi-articulator communication, and study the sources of variation in production and outcomes in the great ape genus that is most suitable for this avenue of research: orang-utans (*Pongo* spp.). First, the orang-utan populations of Borneo (*Pongo pygmaeus wurmbii*) and Northwest-Sumatra (i.e. Suaq and Ketambe, *Pongo abelii*) differ considerably in sociability<sup>26,cf. 27</sup> and social tolerance (Bornean orang-utans become more stressed in group settings than Sumatrans<sup>28</sup>). The consistently higher level of sociability in Sumatrans may lead to a greater need to refine messages conveyed in signals, and thus to more multi-articulator communicative acts. Second, in contrast to natural environments, captive orang-utans are always in close proximity and more on the ground<sup>9,15,19</sup>, thus reducing the need for multi-sensory signals. Their sociability is also not constrained by food availability<sup>29</sup>. In the wild individuals may have fewer interaction opportunities and communication is hampered by arboreality and obscuring vegetation, whereas captivity enables frequent interactions and short-distance communication with conspecifics other than the mother. Third, the pairing of social partners (interaction dyad) also affects features of social interactions irrespective of captive-wild and Bornean-Sumatran contrasts, e.g. due to differences in social tolerance and familiarity<sup>30,31</sup>. Although mothers are the most important communication partner of infant orang-utans<sup>10,32,33</sup>, temporary associations during feeding or travelling occur, particularly if food is abundant<sup>34,35</sup>, thus providing opportunities for social interactions beyond the mother-infant unit<sup>36-38</sup>. We expect that the reduced familiarity of these dyads, and thus the greater uncertainty of tolerance, would lead them to use more multi-articulator signals.

In the present study, we examined close-range communicative interactions of Bornean and Sumatran orang-utans in two wild populations and five zoos. While focal units in this study consisted of mothers and their dependent offspring, we also examined interactions with and among other members of the group/temporary association. By examining species differences related to differential sociability on one hand, and recipient-dependent factors on the other, we aimed to evaluate two major hypotheses explaining multimodal signal function discussed for great apes: redundancy and refinement. Since there are virtually no studies applying a similar comparative approach to any primate species, our predictions are largely exploratory.

If multi-sensory and multi-articulator communicative acts indeed function as ‘backup signals’, we first predicted that that these acts (comprising e.g. visual plus auditory) are more effective (i.e. more likely to result in the apparently intended outcome) than the single (e.g. purely visual) constituent parts, but have little or no effect on the type of outcome (i.e. dominant versus sub-dominant interaction outcome, see below). A second prediction would be that multi-component acts should be more common in the wild than in captive settings, where semi-solitariness limits interaction opportunities and communication is hampered by arboreality and obscuring vegetation<sup>15,19,39</sup>.

If, on the other hand, multi-component acts serve primarily to refine messages, we predicted that they would be used more often for subdominant communicative goals (i.e. reducing ambiguity). For instance, if a certain communicative act is most frequently (> 50%) produced to solicit food transfers, but also in others (e.g. grooming, co-locomote), we predict that this communicative act is accompanied by other constituent parts (e.g. facial expression or gaze) more often in non-begging than begging interactions. Second, we predicted that multi-component acts would be more common in settings and interactions with higher uncertainty and more differentiated interactions in the social environment<sup>12,14,15</sup>. Specifically, we expected an effect of species- and dyad-dependent effect of setting: although wild individuals may use more acts associated with gaze than their captive counterparts (due to lower degrees of social tolerance and thus less predictable outcomes), this effect should be more pronounced in Sumatrans (i.e. the more sociable population) and in interactions beyond the mother-offspring unit.

A secondary aim was to examine the sources of variation in the individual sensory modalities and articulators that constitute multi-component communication in orang-utans. Inevitably, some modalities and articulators are predicted to be involved more in the communication process than others: in natural settings, dense vegetation in the canopy means that there are fewer opportunities for the direct lines-of-sight needed for visual communication. As arboreal species, orang-utans are thus thought to rely less on purely visual signals than other (e.g. audio-visual) communicative means<sup>34,40</sup>.

## Methods

### Study sites

Data were collected at two field sites and five captive facilities (zoos). We observed wild orang-utans at the long-term research sites of *Suaq Balimbing* (03°02'N; 97°25'E, Gunung Leuser National Park, South Aceh, Indonesia) and *Tuanan* (02°15'S; 114°44'E, Mawas Reserve, Central Kalimantan, Indonesia), on a population of wild Sumatran (*Pongo abelii*) and Bornean orang-utans (*Pongo pygmaeus wurmbii*), respectively. Both study sites consist mainly of peat swamp forest and show high orang-utan densities, with 7 individuals per km<sup>2</sup> at *Suaq* and 4 at *Tuanan*<sup>41,42</sup>. Captive Bornean orang-utans were observed at the zoos of Cologne and Munster, and at Apenheul (Apeldoorn), while Sumatran orang-utans were observed at the zoo of Zurich and at Hellabrunn (Munich) (see EEP studbook for details on captive groups<sup>43</sup>). While captive Sumatran orang-utans were housed in groups of nine individuals each, captive

Bornean groups were generally smaller and sometimes included only a mother and her dependent and independent offspring (e.g. Apenheul). Signallers included in this study consisted of 33 Bornean (21 wild/12 captive) and 38 Sumatran orang-utans (20 wild/18 captive; see Tab. S1 for detailed information on subjects and group compositions). Recipients consisted of 81 individuals comprising all age-sex classes.

## Data collection

Focal observations were conducted between November 2017 and October 2018 (Suaq Balimbing: November 2017 – October 2018; Tuanan: January 2018 – July 2018, European zoos: January 2018 – June 2018). At the two field sites, they consisted of full (nest-to-nest) or partial follows (e.g. nest-to-lost or found-to-nest) of mother-infant units, whereas in zoos 6-hour focal follows were conducted. Behavioural data were collected following an established protocol for orang-utan data collection (<https://www.aim.uzh.ch/de/orangutannetwork/sfm.html>), using focal scan sampling. All observers were trained to use this protocol and inter-observer reliability tests were conducted after each training phase. MF collected data at both study sites ensuring the use of the same criteria during training see also <sup>44</sup>. Two different behavioural sampling methods were combined: First, intra-specific communicative interactions of all social interactions of the focal as signaller and as receiver with all partners, but also among other conspecifics present were recorded using a digital High-Definition camera (Panasonic HC-VXF 999 or Canon Legria HF M41) with an external directional microphone (Sennheiser MKE600 or ME66/K6). In captive settings with glass barriers, we also used a Zoom H1 Audio recorder that was placed in background areas of the enclosure whenever possible, to enable the recording of auditory communicative acts. Second, using instantaneous scan sampling at ten-minute intervals, we recorded complementary data on the activity of the focal individual, the distance and identity of all association partners, in case of social interactions the interaction partner as well as several other parameters. In captive settings with glass barriers, we also used a Zoom H1 Audio recorder that was placed in background areas of the enclosure whenever possible, to enable the recording of auditory communicative acts. During ca. 1600 hours of focal observations, we video-recorded more than 6300 communicative interactions which were subsequently screened for good enough quality to ensure video coding.

## Coding procedure

A total of 2655 high-quality recordings of orang-utan interactions (wild: 1643, captive: 1012), which could each include multiple communicative acts, were coded using the program BORIS version 7.0.4. <sup>45</sup>. We designed a coding scheme to enable the analysis of articulators and sensory modalities involved in presumably communicative acts directed at conspecifics (i.e. close-range social behaviours that apparently served to elicit a behavioural change in the recipient and were mechanically ineffective; see also Table 1) and thus included “potential gestures” <sup>46</sup>. Cases where acts were directed at observers or achieved their presumed goal *sensu* <sup>46</sup> (apparent aim of the signaller based on the individuals involved and the immediate social context) via physical force (e.g. nursing solicitations, infant collections) were thus excluded from the dataset. For each communicative act, we coded the following “modifiers”: body parts involved in production (hands, head, torso), sensory modalities involved in perception (e.g. visual,

tactile), presumed goal (e.g. share food/object, play/affiliate, co-locomotion, following the distinctions of <sup>46</sup>), gaze direction (e.g. recipient, object), recipient’s attentional state (e.g. faced towards signaller, faced away), interaction outcome (e.g. apparently satisfactory outcome) and other variables not relevant in this study (see Tab. S2 for levels and definitions of all coded variables). We distinguished multi-articulator (i.e. at least two different articulators simultaneously involved in production) from multi-sensory communicative acts (i.e. at least two salient sensory modalities simultaneously involved in perception, see Table 1). While the specific communicative acts comprising repertoires are out of the scope of this paper and part of different analyses <sup>47</sup>, we used this dataset to determine the dominant outcome of a communicative act in a given research setting and orang-utan species (see ESM\_3\_outcomes).

During an initial training period, and afterwards in regular intervals, coding performance of all observers was evaluated with alternating datasets using the Cohen’s Kappa coefficient to ensure inter-coder reliability <sup>48</sup>. Trained coders (MF, NB, CF, CW, LM, MJ) proceeded with video coding only if at least a ‘good’ level ( $\kappa = 0.75$ ) of agreement was found for type of communicative act, articulator, sensory modality, presumed goal, and outcome. For further details on the distribution of coded interactions across species, settings and interaction dyads see Table S3.

Table 1  
Definition and operationalisation of relevant terms used in analyses

Term	Definition	Coding in GLMMs
Communicative act (CA)	Close-range socially directed, non-coercive behaviour eliciting a behavioural change in another individual	Rows of dataset
Articulator	Body part involved in production of CAs, such as manual, bodily, gaze, facial and vocal acts	each 2 levels, e.g. not manual (0), manual (1)
Sensory channel	Modality involved in perception of CAs, such as visual, tactile, auditory, seismic	each 2 levels, e.g. not visual (0), visual (1)
Multi-sensory acts	CAs perceived through at least two different sensory channels (e.g. visual, tactile, auditory)	2 levels: uni-sensory (0) / multi-sensory (1)
Multi-articulator acts	CAs involving at least two different articulators (e.g. limbs, gaze, voice)	2 levels: uni-articulatory (0) / multi-articulatory (1)
Effectiveness	Presence or absence of “apparently satisfactory outcome” (ASO) <i>sensu</i> <sup>11</sup>	2 levels: other than ASO (0), ASO (1)
Dominant outcome match	Presumed goal <i>sensu</i> <sup>46</sup> matches interaction outcome that is most commonly (> 50%) attributed to a specific CA <i>sensu</i> <sup>47</sup>	2 levels: sub-dominant (0) / dominant (1)

## Statistical analyses

For the dataset of around 7578 communicative acts resulting from the coding procedure, we used Generalized Linear Mixed Models<sup>49</sup> with a binomial error structure and logit link function. We investigated sources of variation in four sets of response variables, referring to (a) employment of sensory modalities (visual, tactile, auditory, seismic, multi-sensory), (b) employment of articulators (manual, bodily, facial, vocal, gaze, multi-articulatory), (c) effectiveness (i.e. whether or not the signaller received an apparently satisfactory response; *sensu*<sup>11,46</sup>), and (d) dominant outcome match (i.e. whether or not the presumed goal *sensu*<sup>46</sup> of a communicative act matched the major interaction outcome associated with it; see Table 1 for definitions, and ESM\_3\_outcomes for all apparently satisfactory outcomes and the dominant outcome for every type of communicative act for every setting and species<sup>47</sup>).

In models (a) and (b), we included orang-utan species (2 levels: Bornean, Sumatran), research setting (2 levels: captive, wild), kin relationship (3 levels: mother-infant [i.e. only including unweaned immatures], maternal kin, unrelated), and age difference (3 levels: younger age class, same age class, older age class; age classes divided in four levels: infant = dependent unweaned immature, est. 0–5 yrs, juvenile = dependent weaned immature est. 5–8 yrs, adolescent = independent immature, est. 9–14 yrs, adult: sexually active, est. >15 yrs;<sup>50</sup>) as our key test predictors. Because we assumed that the effect of research setting might depend on genetic predisposition (i.e. species), we included the interaction between these two variables into all models. To adjust for multiple comparisons, we tested interaction effects using pairwise contrasts with the function *lsmeans* (with argument “adjust” set to “sidak”) of the package *lsmeans*.

In the models (c) and (d), we included composition of multi-sensory (2 levels: visual/tactile only, visual/tactile plus other modality) or multi-articulator (manual/bodily/gaze only, manual/bodily/gaze plus other articulator) acts as only key test predictor. Communicative acts involving vocal, seismic, facial or vocal components were not common enough to allow inferential analyses (see Table 2). In these analyses, we considered only single-articulator communicative acts when testing the function multi-sensory (e.g. tactile alone versus tactile plus other) and only uni-sensory acts when testing multi-articulatory production. We note that the same cases of multi-sensory acts (e.g. visual plus tactile) could be included in both the visual multi-sensory and tactile multi-sensory models (*vice versa* for multi-articulatory acts).

As great ape dyadic interactions are also profoundly shaped by individual and social variables<sup>8,12</sup>, we included the following variables as further fixed effects (control predictors) into the models: subjects’ age class (3 levels: adult, older immature > 5 years of age, young immature < 5 years of age), sex (2 levels: female, male), and presumed goal (3 levels: share food/object, play/affiliate, other; as most orang-utan close-range interactions in captive settings relate to play or feeding;<sup>51</sup>). In models (c) and (d), species, setting, kin relationship and age difference (see above) were included as control predictors rather than key test predictors as in models (a) and (b). To control for repeated measurements, the identity of the dyad, the subject and the partner were treated as random effects. We further included communicative act

type (e.g. touch, raise limb etc., <sup>47</sup>), group identity, coder identity and video file number (accounting for the fact that communicative acts of the same interaction are non-independent) as random effects. To keep type 1 error rates at the nominal level of 5%, we also included the relevant random slopes components within subject, recipient and dyad ID <sup>52</sup>. All models were implemented in R (v3.4.1, <sup>53</sup>) using the function *glmer* of the package *lme4* <sup>54</sup>. To test the overall significance of our key test predictors <sup>55</sup>, we compared the full models with the respective null models comprising only the control predictors as well as all random effects using a likelihood ratio test <sup>56</sup>. To control for collinearity, we determined the Variance Inflation Factors VIF; <sup>57</sup> from a model including only the fixed main effects using the function *vif* of the package *car* <sup>58</sup>. This revealed collinearity to not be an issue (maximum VIF = 2.4). Tests of the individual fixed effects were derived using likelihood ratio tests (R function *drop1* with argument “test” set to “Chisq”).

## Results

### Production of multi-sensory acts and constituent parts

Out of the 7578 coded communicative acts, 3465 were uni-sensory and uni-articulatory, 1774 multi-articulatory but uni-sensory, 1489 multi-sensory but uni-articulatory, and 859 both multi-articulatory and multi-sensory (see Fig. 1 for Venn diagram). On average, communicative acts were multi-sensory in 25% of observed cases. For specific sensory channels, we found that communicative acts were purely visual in 23% of cases, contained salient tactile components in 75%, salient auditory components in 3%, and salient seismic components in 0.9%. (for detailed results in relation to species and setting, see Table 2; for sources of variation in the employment of specific modalities, see Tab. S4).

Table 2

Mean percentage and SD (%) of communicative acts involving specific modalities and articulators and their outcomes in relation to research setting and orang-utan species.

	Captivity				Wild				Total	
	Bornean		Sumatran		Bornean		Sumatran		Mean	SD
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
<i>Modalities</i>										
Visual	27.4	9.2	31.3	7.6	6.6	8.3	30.9	22.7	<b>23.2</b>	<b>17.5</b>
Tactile	72.0	9.2	66.7	8.4	92.9	8.4	63.5	23.0	<b>74.5</b>	<b>18.6</b>
Auditory	3.2	8.2	1.8	3.2	5.5	13.5	3.8	5.7	<b>3.2</b>	<b>8.2</b>
Seismic	0.0	0.0	0.4	0.7	0.6	1.2	2.3	4.8	<b>0.9</b>	<b>2.8</b>
Multi-sensory	4.3	3.6	19.2	7.3	36.1	16.8	32.4	17.1	<b>25.4</b>	<b>17.6</b>
<i>Articulators</i>										
Manual	71.6	10.1	57.6	13.2	69.5	17.6	68.1	20.8	<b>66.4</b>	<b>17.1</b>
Bodily	23.9	11.7	35.6	13.9	22.4	15.3	16.8	18.7	<b>24.4</b>	<b>16.8</b>
Vocal	0.3	0.8	2.0	3.7	5.2	12.9	2.7	3.8	<b>2.9</b>	<b>7.6</b>
Facial	1.5	2.7	4.6	4.1	0.6	1.2	0.8	1.1	<b>1.8</b>	<b>3.0</b>
Gaze	18.8	11.1	63.3	12.7	68.5	20.8	75.8	17.7	<b>60.8</b>	<b>25.6</b>
Multi-articulatory	6.6	5.2	35.5	10.2	29.8	15.4	42.4	10.3	<b>30.9</b>	<b>16.5</b>
<i>Effectiveness</i>	57.6	16.3	50.9	11.8	72.1	11.5	49.5	13.6	<b>57.9</b>	<b>16.0</b>
<i>Dominant outcome match</i>	69.3	21.6	67.8	23.7	81.5	14.1	65.5	28.3	<b>71.5</b>	<b>22.9</b>

When we tested sources of variation in multi-sensory acts, the full model including the key test predictors (i.e. species x setting, kin relationship, age difference) fitted the data better than the null model (Likelihood ratio tests [LRT] for multi-sensory acts:  $\chi^2_7 = 45.785$ ,  $P < 0.001$ ,  $N = 7587$ ). We found a significant interaction between orang-utan species and research setting (estimate  $\pm$  s.e. =  $-2.032 \pm 0.347$ ,  $\chi^2_1 = 19.691$ ,  $P < 0.001$ ): post hoc Sidak tests showed that wild orang-utans of both species employed more multi-sensory acts than captive ones (Borneans:  $-2.864 \pm 0.299$ ,  $Z = -9.592$ ,  $P < 0.001$ , Sumatrans:  $-0.832 \pm 0.181$ ,  $Z = -4.587$ ,  $P < 0.001$ ), and that captive Sumatrans produced more multi-sensory acts than captive Borneans ( $-1.73 \pm 0.298$ ,  $Z = -5.817$ ,  $P < 0.001$ ; Fig. 2). In addition, multi-sensory acts were significantly less likely in interaction with maternal kin ( $-0.427 \pm 0.162$ ,  $\chi^2_1 = 6.885$ ,  $P = 0.009$ ). For effects of non-significant key predictors and those of the control variables see Table S4.

# Production of multi-articulatory acts and constituent parts

On average, communicative acts comprised combinations of different articulators (i.e. multi-articulator acts) in 31% of instances. In terms of specific articulators, communicative acts involved manual acts in 66% of observed cases, bodily acts in 24%, facial acts in 2%, vocal acts in 3% and gaze in 61% of cases (for detailed results in relation to species and setting, see Table 2; for sources of variation in the employment of specific articulators, see Tab. S4). Overall, the full model including the key test predictors fitted the data better than the null model (LRT for multi-articulator acts:  $\chi^2_7 = 52.83$ ,  $P < 0.001$ ,  $N = 7587$ ). Specifically, we found a significant interaction between species and setting ( $-1.521 \pm 0.336$ ,  $\chi^2_1 = 12.775$ ,  $P < 0.001$ ; Fig. 3): post hoc Sidak tests showed that Sumatran orang-utans in either research setting employed more multi-articulator acts than Bornean ones (captivity:  $-2.196 \pm 0.285$ ,  $Z = -7.717$ ,  $P < 0.001$ , wild:  $-0.675 \pm 0.184$ ,  $Z = -3.664$ ,  $P < 0.001$ ), and that wild orang-utans of both species employed more multi-sensory acts than captive ones (Borneans:  $-2.224 \pm 0.293$ ,  $Z = -7.591$ ,  $P < 0.001$ , Sumatrans:  $-0.703 \pm 0.176$ ,  $Z = -4.002$ ,  $P < 0.001$ ). Communicative acts among mother-infant interactions ( $-0.86 \pm 0.175$ ,  $\chi^2_1 = 21.716$ ,  $P < 0.001$ ; Fig. 3) were less likely to be multi-articulator than those among other interaction dyads. For effects of non-significant key predictors and those of control variables see Table S4.

## Effectiveness of multi-component acts

Communicative acts received apparently satisfactory responses in 58% of observed cases (for detailed results in relation to species and setting see Table 2). We tested whether the uni- versus multi-component use of different modalities (i.e. visual/tactile act involved) and articulators (i.e. bodily/manual act/gaze involved) predicted the probability of receiving an apparently satisfactory outcome. The full models including the key test predictor (i.e. multi-component act) fitted the data better than the null models for combinations of sensory modalities (LRT visual multi-sensory:  $\chi^2_1 = 14.18$ ,  $P < 0.001$ ,  $N = 2301$ , see Fig. 4a; tactile multi-sensory:  $\chi^2_1 = 10.085$ ,  $P < 0.001$ ,  $N = 3743$ , see Fig. 4b), as well as for multi-component use of bodily acts and gaze (LRT bodily multi-articulator:  $\chi^2_1 = 4.288$ ,  $P = 0.038$ ,  $N = 1498$ , gaze multi-articulator:  $\chi^2_1 = 12.813$ ,  $P < 0.001$ ,  $N = 4513$ ). See Table S5 for effects of non-significant key predictors and those of control variables. Thus, both (uni-articulator) multi-sensory and (uni-sensory) multi-articulator communicative acts were significantly more likely to result in appropriate interaction outcomes than their isolated constituent parts, irrespective of setting, species or communicative act.

## Association with dominant outcomes by multi-component acts

Communicative acts were associated with their dominant outcomes in 72% of observed cases (for detailed results in relation to species and setting see Table 2). We tested whether the multi-component use of different modalities (i.e. visual multi-sensory, tactile multi-sensory) and articulators (i.e. manual multi-articulator, bodily multi-articulator, gaze multi-articulator) predicted the likelihood of a communicative act matching the predominant outcome of a specific type of communicative act. Our key test predictor (i.e. isolated versus combined use) significantly enhanced the model fit for multi-articulator

acts except for those involving manual acts (LRT bodily multi-articulator:  $\chi^2_1 = 6.664$ ,  $P = 0.01$ ,  $N = 1429$ , see Fig. 5a; gaze multi-articulator:  $\chi^2_1 = 6.702$ ,  $P = 0.01$ ,  $N = 3869$ , see Fig. 5b; manual multi-articulator:  $\chi^2_1 = 1.101$ ,  $P = 0.294$ ,  $N = 2590$ ). See Table S6 for effects of non-significant key predictors and those of control variables. Thus, uni-sensory communicative acts consisting of at least two articulators (e.g. gaze plus bodily act) were significantly less likely to match dominant interaction outcomes than their isolated (e.g. bodily act alone) constituent parts, irrespective of setting, species or type of communicative act. No such effect was found for multi-sensory acts (LRT tactile multi-sensory:  $\chi^2_1 = 2.658$ ,  $P = 0.103$ ,  $N = 3129$ ).

## Discussion

This study was aimed at disentangling multi-sensory and multi-articulator communication, and at deciphering the constituting elements (that is, specific sensory modalities and articulators, respectively) in wild and captive orang-utans of two different species. Moreover, by studying the effects of species and research setting on signallers' behaviour, as well as effects of the employment of multi-component acts on responses and types of interaction outcomes, we wanted to gain insight into the functions of these two forms of multi-component communication.

One key finding of this study is that both multi-sensory and multi-articulator differ from the respective uni-component acts in both production and effects, but may still have different functions depending on social circumstances. Thus, we can greatly improve our understanding of the function of "multimodality" if we tease apart the articulators and modalities involved. We will first attend to our predictions and results regarding multi-sensory communication.

When controlling for the type of communicative act, multi-sensory (uni-articulator) acts involving both visual and tactile components were more likely to receive apparently satisfactory responses (i.e. outcomes that resulted in the cessation of communication *sensu*<sup>11</sup>, and that thus matched the presumed goals *sensu*<sup>46</sup>) than their uni-sensory constituent parts. In contrast, we found no evidence that multi-sensory acts predicted sub-dominant interaction outcomes. Therefore, in orang-utans, multimodality *sensu stricto* seems to primarily enhance effectiveness rather than reducing ambiguity: communication through multiple sensory channels in orang-utans may facilitate detection and thereby effective communication by increasing the reliability of the message content, consistent with a redundancy function<sup>18</sup>. These findings have to be viewed with caution given that multi-sensory communication in our study mainly involved visual and tactile (rather than auditory and seismic components) in close-range interactions, and that we probably missed some low-amplitude auditory acts (e.g. vocalizations) due to environmental constraints (e.g. glass barriers in captivity or noisy surroundings in the field). However, the gestural repertoire of great apes has indeed been considered to be widely redundant<sup>59</sup>, and studies conducted in communities of wild chimpanzees<sup>60,61</sup> suggest that both simultaneously and sequentially redundant signalling might play a particular role in certain developmental stages in apes, as a mechanism to learn context-appropriate communicative techniques termed 'repertoire tuning' by<sup>61</sup>.

In line with our expectation that the arboreal setting would impose particular communicative affordances, we found that multi-sensory communication was more commonly observed in wild than captive orang-utans, although this difference was more pronounced for the Bornean species. Accordingly, for orang-utans the benefits of communicating in several sensory channels at once (as a “back-up strategy”) at the expense of subtler communicative acts may be greater in the wild, where greater competition due to food scarcity may require facilitation of mutual understanding, and particularly among Bornean orang-utans, where they are rarer (both in captivity and in the wild).

With regard to the predictions concerning the signaller-based (multi-articulator) perspective, our results suggest that multi-articulator communication may serve, at least under certain circumstances (e.g. those communicative acts that are purely visual), to reduce ambiguity *in addition to* boosting effectiveness. First and as predicted, we found that multi-articulator acts, as well as non-tactile acts and those involving gaze, were more common in dyads other than mother-infant, but also in Sumatran orang-utans in either research setting irrespective of interaction dyad. The profound difference between mother-offspring and other interactions is arguably due to the high trust the signaller can have that the recipient is socially tolerant. This finding corroborates previous work on wild chimpanzees, demonstrating that purely visual, non-contact communication may be more relevant in interactions with less socially tolerant conspecifics<sup>30,31</sup>. Although orang-utans tend to have fewer opportunities for social interactions than the African apes outside the mother-offspring bond (but see<sup>62</sup> showing overlap in ‘solitariness’ between eastern chimpanzee females and North-West-Sumatran orang-utans), they do occur regularly. This is especially true in the populations of North-West Sumatra due to frequent encounters with adult males<sup>26,63</sup>, but also in some Bornean populations for mother-infant pairs of larger matrilineal clusters<sup>36,37</sup>. The environments that captive and wild Sumatran orang-utans inhabit, at least in this study, were characterized by more frequent encounters with conspecifics and a thus probably a wider set of possible social partners see also<sup>37</sup>. Social interactions with conspecifics beyond the matriline are rarer in Tuanan than in Suaq<sup>37</sup>, as are unpredictable outcomes of interactions that would require subtler communication from a larger distance. Taken together, our results strikingly demonstrate that orang-utan signallers are able to flexibly adjust their signalling to specific recipients, in line with previous work on African apes e.g.<sup>31,64</sup>.

Second, and again in line with our prediction, multi-articulator uni-sensory acts (i.e. bodily or gaze accompanied by other means of the same sensory modality) were more likely to be produced when the presumed goal of the interaction did *not* match with the dominant interaction outcome of a particular communicative act. This finding strongly suggests that constituent parts of multi-articulator acts are non-redundant and thus, may serve to refine the message<sup>15</sup>. Human and ape communication have in common that signals are not always tightly coupled with a given referent: meaning does not only depend on the communicative act that is being used but also on the interaction history, contextual information and social aspects of the interaction<sup>59,65,66</sup>. Importantly, by combining articulators in social interactions, interactants are able to clarify their ambiguous but thereby efficient “main signals” (e.g. speech acts in humans<sup>6,23</sup>). Nevertheless, our results also demonstrate that multi-articulator uni-sensory acts are more effective than their uni-articulator constituent parts. It is probably not surprising that successful

disambiguation also results in more appropriate responses, which suggests that effectiveness alone is not sufficient to disentangle hypotheses for the function of multi-component communication. Nonetheless, our findings are consistent with the notion that the redundancy function applies more to multi-sensory signalling, whereas refinement applies more to multi-articulator signalling.

Importantly, multi-articulator communication in orang-utan consisted mainly of manual/bodily acts (“potential gestures” *sensu*<sup>46</sup>) associated with eye gaze (constituting “gestures” according to common definitions in comparative research), rather than with vocalizations or facial expressions. Multi-articulator acts involving vocalizations were rare (161 out of 7578 acts in total), which was probably largely due to the overall rare use of vocalization in orang-utan close-range communication<sup>10</sup>, but is also consistent with reports of relatively rare use of gesture-vocal combinations in chimpanzees<sup>12,13</sup> and bonobos<sup>14,25</sup>. It is important to note that gaze, even though it definitely has a communicative function, often acts as a social cue rather than an intentionally produced signal. However, we do know that orang-utans are capable of controlling their gaze for bouts of intentional communication<sup>32,33,67</sup>, suggesting that eye gaze serves as important communicative articulator just as it does in humans. As important component of social interactions, gaze can be directed at specific individuals (thereby being less ambiguous than auditory and olfactory signals), and may be used to predict another individual's behaviour<sup>68</sup>. We speculate that unrelated orang-utans are generally much more unpredictable in their responses, so they may have evolved a strong tendency to visually check the emotional state of their potential interaction partners.

Because previous studies on great apes mainly focused on multi-articulator communication (and specifically the function of gesture-vocal combinations), it is unclear to what extent communicative acts were actually also multi-sensory (i.e. silent gestures plus audible vocalization). Captive bonobos, but not chimpanzees, have been shown to be more responsive to multi-articulator (i.e. gestures combined with facial/vocal signals; called multi-modal by the authors) than uni-articulator communication despite its rare usage<sup>25</sup>. Moreover, male bonobos use the same vocalization (“contest-hoot”) in playful and aggressive contexts but add gestures to distinguish between the two<sup>14</sup>. For wild chimpanzees, responses to combinations of gestures and vocalizations were more likely to match the response of the gestural than the vocal components<sup>13</sup>. In line with this study, wild chimpanzees, after perceived goals were not achieved, switched to gesture-vocalization combinations only if the initially single signals were vocal<sup>12</sup>. Thus, evidence so far, including our own work, suggests that the combination of different articulators in great ape communication is apparently non-redundant, and serves to resolve ambiguity in the communicative act.

In sum, this study's aim was to disentangle multi-articulator and multi-sensory communication in great ape communication, and study their respective functions, by focusing on communicative interactions within and outside the highly tolerant mother-infant bond. On one hand, we showed that multi-sensory acts (e.g. visual plus tactile) in orang-utans seem to have an important redundancy function as these more often received more apparently satisfactory responses (i.e. were more effective) than corresponding

uni-sensory acts and were more common in wild populations. On the other hand, multi-articulator acts (which consisted to a large part of manual/bodily acts accompanied by gaze) were more common for communicative acts whose presumed goal did not match with the dominant outcome, suggesting that gaze seems to have a refining (disambiguating) function in these instances. In addition, multi-articulatory communication was apparently more relevant in interactions with less socially tolerant interaction partners where outcomes are less predictable, requiring the usage of refining (or disambiguating) acts. This is consistent with human communication, in which multi-sensory (audio-visual) messages were shown to be processed faster, and gestural and facial acts accompanying spoken language serve to refine and disambiguate the message conveyed in speech acts<sup>6,23</sup>.

The finding that functions of multi-sensory and multi-articulatory communication may differ depending on the specific sensory modalities and articulators involved demonstrates the importance of empirically distinguishing between these forms of communication. It is therefore all the more important that comparative studies do not compare apples with oranges: the increase of multi-component study designs in primate communication is timely, but comparisons with human communication will be most fruitful if the difference between *production* and *perception* of communicative acts is explicitly addressed. Implementing such a biological meaningful comparative approach to non-human species will comprise an invaluable tool to study the origins of the human multi-component communication system.

## Declarations

### Acknowledgements

We thank Caroline Schuppli (Suaq), Erin Vogel and Suci Utami Atmoko (Tuanan), Kerstin Bartesch (Tierpark Hellabrunn), Claudia Rudolf von Rohr (Zoo Zürich), Alexander Sliwa (Kölner Zoo), Simone Sheka (Allwetterzoo Münster) and Thomas Bionda (Apenheul Primate Park) as well as all research staff and zoo keepers for a fruitful collaboration during this study. We gratefully acknowledge Clemens Becker for providing the EEP studbook, the Indonesian State Ministry for Research and Technology (RISTEK), the Indonesian Institute of Science (LIPI), the Directorate General of Natural Resources and Ecosystem Conservation – Ministry of Environment & Forestry of Indonesia (KSDAE-KLHK), the Ministry of Internal affairs, the Nature Conservation Agency of Central Kalimantan (BKSDA), the local governments in Central Kalimantan, the Kapuas Protection Forest Management Unit (KPHL), the Bornean Orang-utan Survival Foundation (BOSF) and MAWAS in Palangkaraya. Moreover, we thank Simone Pika and Eva Luef for providing some of the essential equipment, Santhosh Totagera for coding assistance, Uli Knief for statistical advice and Alexander Hausmann for providing a customized jitter-plot function.

### Funding

MF was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, grant FR 3986/1-1) the Forschungskredit Postdoc of the University of Zurich (grant FK-17-106), and the

Stiftung Mensch und Tier (Freiburg). CvS acknowledges the support of the NCCR Evolving Language Program (SNF #51NF40\_180888).

### Authors' contributions

MF and CvS conceived of the study. MF designed the project, collected, coded and analysed data. NB, CF, CW, LM, MJ helped to collect, curate and code data. TMS, MvN and CvS provided resources. MF wrote the manuscript with critical inputs from MvN and CvS. All authors approved the submission of the manuscript.

### Competing interests

The authors declare that they have no conflict of interest.

### Data availability

The datasets supporting this article (ESM\_2\_data and ESM\_3\_outcomes) are available on GitHub: <https://github.com/MarlenF/multi-orang>.

### Code availability

The R code supporting this article (ESM\_4\_Rcode and ESM\_5\_Rcode) are available on GitHub: <https://github.com/MarlenF/multi-orang>.

## References

1. Goldin-Meadow, S. The role of gesture in communication and thinking. *Trends in Cognitive Sciences* **3**, 419–429 (1999).
2. McNeill, D. *Language and gesture*. (Cambridge University Press, 2000).
3. Holle, H., Gunter, T. C., Rüschemeyer, S.-A., Hennenlotter, A. & Iacoboni, M. Neural correlates of the processing of co-speech gestures. *NeuroImage* **39**, 2010–2024 (2008).
4. Holler, J., Kendrick, K. H. & Levinson, S. C. Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic Bulletin & Review* **25**, 1900–1908 (2018).
5. Stivers, T. *et al.* Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* **106**, 10587–10592 (2009).
6. Holler, J. & Levinson, S. C. Multimodal language processing in human communication. *Trends in Cognitive Sciences* **23**, 639–652 (2019).
7. Higham, J. P. & Hebets, E. A. An introduction to multimodal communication. *Behav. Ecol. Sociobiol.* **67**, 1381–1388 (2013).
8. Liebal, K., Waller, B. M., Burrows, A. M. & Slocombe, K. E. *Primate Communication: A Multimodal Approach*. (Cambridge University Press, 2013).

9. Partan, S. R. & Marler, P. Communication goes multimodal. *Science* **283**, 1272–1273 (1999).
10. Fröhlich, M., Lee, K., Mitra Setia, T., Schuppli, C. & van Schaik, C. P. The loud scratch: a newly identified gesture of Sumatran orangutan mothers in the wild. *Biol. Lett.* **15**, 20190209 (2019).
11. Hobaiter, C. & Byrne, Richard W. The meanings of chimpanzee gestures. *Curr. Biol.* **24**, 1596–1600 (2014).
12. Hobaiter, C., Byrne, R. W. & Zuberbühler, K. Wild chimpanzees' use of single and combined vocal and gestural signals. *Behav. Ecol. Sociobiol.* **71**, 96 (2017).
13. Wilke, C. *et al.* Production of and responses to unimodal and multimodal signals in wild chimpanzees, *Pan troglodytes schweinfurthii*. *Anim. Behav.* **123**, 305–316 (2017).
14. Genty, E., Clay, Z., Hobaiter, C. & Zuberbühler, K. Multi-modal use of a socially directed call in bonobos. *PLoS ONE* **9**, e84738 (2014).
15. Fröhlich, M. & van Schaik, C. P. The function of primate multimodal communication. *Anim. Cogn.* **21**, 619–629 (2018).
16. Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T. & van Schaik, C. P. Multimodal communication and language origins: integrating gestures and vocalizations. *Biol Rev* **94**, 1809–1829 (2019).
17. Hage, S. R. & Nieder, A. Dual neural network model for the evolution of speech and language. *Trends in neurosciences* **39**, 813–829 (2016).
18. Johnstone, R. A. Multiple displays in animal communication: 'backup signals' and 'multiple messages'. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **351**, 329–338 (1996).
19. Hebets, E. A. & Papaj, D. R. Complex signal function: Developing a framework of testable hypotheses. *Behavioural Ecology and Sociobiology* **57**, 197–214 (2005).
20. Özyürek, A. Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130296 (2014).
21. Holler, J. & Wilkin, K. Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and cognitive processes* **24**, 267–289 (2009).
22. Holler, J. *et al.* Eye'm talking to you: speakers' gaze direction modulates co-speech gesture processing in the right MTG. *Social Cognitive and Affective Neuroscience* **10**, 255–261 (2014).
23. Holle, H. & Gunter, T. C. The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of cognitive neuroscience* **19**, 1175–1192 (2007).
24. McGurk, H. & MacDonald, J. Hearing lips and seeing voices. *Nature* **264**, 746–748 (1976).
25. Pollick, A. S. & de Waal, F. B. M. Ape gestures and language evolution. *Proceedings of the National Academy of Sciences* **104**, 8184–8189 (2007).
26. van Schaik, C. P. The socioecology of fission-fusion sociality in orangutans. *Primates* **40**, 69–86 (1999).
27. Roth, T. S., Rianti, P., Fredriksson, G. M., Wich, S. A. & Nowak, M. G. Grouping behavior of Sumatran orangutans (*Pongo abelii*) and Tapanuli orangutans (*Pongo tapanuliensis*) living in forest with low

- fruit abundance. *Am. J. Primatol.* **n/a**, e23123 (2020).
28. Weingrill, T., Willems, E. P., Zimmermann, N., Steinmetz, H. & Heistermann, M. Species-specific patterns in fecal glucocorticoid and androgen levels in zoo-living orangutans (*Pongo* spp.). *Gen. Comp. Endocrinol.* **172**, 446–457 (2011).
  29. Maple, T. L. *Orangutan behavior*. (Van Nostrand Reinhold Co., 1980).
  30. Fröhlich, M., Müller, G., Zeiträg, C., Wittig, R. M. & Pika, S. Begging and social tolerance: Food solicitation tactics in young chimpanzees (*Pan troglodytes*) in the wild. *Evolution and Human Behavior* **41**, 126–135 (2020).
  31. Fröhlich, M., Wittig, R. M. & Pika, S. Play-solicitation gestures in chimpanzees in the wild: flexible adjustment to social circumstances and individual matrices. *R Soc Open Sci* **3**, 160278 (2016).
  32. Bard, K. A. Intentional behavior and intentional communication in young free-ranging orangutans. *Child Dev.* **63**, 1186–1197 (1992).
  33. Knox, A. *et al.* Gesture use in communication between mothers and offspring in wild orang-utans (*Pongo pygmaeus wurmbii*) from the Sabangau Peat-Swamp Forest, Borneo. *International Journal of Primatology* **40**, 393–416 (2019).
  34. MacKinnon, J. The behaviour and ecology of wild orang-utans (*Pongo pygmaeus*). *Anim. Behav.* **22**, 3–74 (1974).
  35. Sugardjito, J., Te Boekhorst, I. & Van Hooff, J. Ecological constraints on the grouping of wild orang-utans (*Pongo pygmaeus*) in the Gunung Leuser National Park, Sumatra, Indonesia. *International Journal of Primatology* **8**, 17–41 (1987).
  36. van Noordwijk, M. A. *et al.* Female philopatry and its social benefits among Bornean orangutans. *Behav. Ecol. Sociobiol.* **66**, 823–834 (2012).
  37. Fröhlich, M. *et al.* Social interactions and interaction partners in infant orang-utans of two wild populations. *Anim. Behav.* (in press).
  38. Schuppli, C. *et al.* The effects of sociability on exploratory tendency and innovation repertoires in wild Sumatran and Bornean orangutans. *Sci. Rep.* **7**, 1–12 (2017).
  39. Partan, S. R. & Marler, P. Issues in the classification of multimodal communication signals. *American Naturalist* **166**, 231–245 (2005).
  40. Rijksen, H. D. *A fieldstudy on Sumatran orang utans (Pongo pygmaeus abelii, Lesson 1827): Ecology, behaviour and conservation*. (H. Veenman 1978).
  41. Singleton, I., Knott, C., Morrogh-Bernard, H., Wich, S. & van Schaik, C. in *Orangutans: Geographic variation in behavioral ecology and conservation* (eds S. A. Wich, S. S. Utami-Atmoko, T Mitra Setia, & C. P. van Schaik) 205–212 (Oxford University Press, 2009).
  42. Husson, S. J. *et al.* in *Orangutans: Geographic variation in behavioral ecology and conservation* (eds S. A. Wich, S. S. Utami-Atmoko, T Mitra Setia, & C. P. van Schaik) 77–96 (Oxford University Press, 2009).
  43. Becker, C. EEP studbook for zoo-housed orang-utans. (2016).

44. Fröhlich, M. *et al.* Social interactions and interaction partners in infant orang-utans of two wild populations. *Anim. Behav.* **166**, 183–191 (2020).
45. Friard, O. & Gamba, M. BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution* **7**, 1325–1330 (2016).
46. Cartmill, E. & Byrne, R. Semantics of primate gestures: intentional meanings of orangutan gestures. *Anim. Cogn.* **13**, 793–804 (2010).
47. Fröhlich, M. *et al.* Wild-captive contrasts in communicative repertoires and functional specificity in orang-utans. (submitted).
48. Bakeman, R. & Quera, V. *Sequential analysis and observational methods for the behavioral sciences.* (Cambridge University Press, 2011).
49. Baayen, R. H. *Analyzing linguistic data.* (Cambridge University Press, 2008).
50. van Noordwijk, M. *et al.* in *Orangutans: geographic variation in behavioral ecology and conservation* (eds SA Wich, SS Utami Atmoko, T Mitra Setia, & CP van Schaik) 189–203 (Oxford University Press, 2009).
51. Liebal, K., Pika, S. & Tomasello, M. Gestural communication of orangutans (*Pongo pygmaeus*). *Gesture* **6**, 1–38 (2006).
52. Schielzeth, H. & Forstmeier, W. Conclusions beyond support: overconfident estimates in mixed models. *Behav. Ecol.* **20**, 416–420 (2009).
53. R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2017).
54. Bates, D., Maechler, M., Bolker, B. & Walker, S. lme4: Linear mixed-effects models using Eigen and S4. *R package version 1.1-7.* (2014).
55. Forstmeier, W. & Schielzeth, H. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behav. Ecol. Sociobiol.* **65**, 47–55 (2011).
56. Dobson, A. J. *An Introduction to Generalized Linear Models.* (Chapman & Hall/CRC, 2002).
57. Quinn, G. P. & Keough, M. J. *Experimental design and data analysis for biologists.* (Cambridge University Press, 2002).
58. Fox, J. & Weisberg, S. *An R companion to applied regression.* 2 edn, (Sage, 2011).
59. Byrne, R. W. *et al.* Great ape gestures: intentional communication with a rich set of innate signals. *Anim. Cogn.* **20**, 755–769 (2017).
60. Fröhlich, M., Wittig, R. M. & Pika, S. Should I stay or should I go? Initiation of joint travel in mother–infant dyads of two chimpanzee communities in the wild. *Anim. Cogn.* **19**, 483–500 (2016).
61. Hobaiter, C. & Byrne, R. W. Serial gesturing by wild chimpanzees: Its nature and function for communication. *Anim. Cogn.* **14**, 827–838 (2011).
62. Wich, S. A., Sterck, E. H. & Utami Atmoko, S. S. Are orang-utan females as solitary as chimpanzee females? *Folia Primatol.* **70**, 23 (1999).

63. Mitra Setia, T., Delgado, R., Utami Atmoko, S., Singleton, I. & van Schaik, C. P. in *Orangutans: Geographic variation in behavioral ecology and conservation* (eds Serge A. Wich, SS Utami Atmoko, T Mitra Setia, & Carel P van Schaik) 245–253 (Oxford University Press, 2009).
64. Genty, E., Neumann, C. & Zuberbühler, K. Bonobos modify communication signals according to recipient familiarity. *Sci. Rep.* **5**, 16442 (2015).
65. Clark, H. H. *Using language*. (Cambridge University Press, 1996).
66. Piantadosi, S. T., Tily, H. & Gibson, E. The communicative function of ambiguity in language. *Cognition* **122**, 280–291 (2012).
67. Rossano, F. & Liebal, K. Requests” and “offers” in orangutans and human infants. *Requesting in social interaction*, 333–362 (2014).
68. Emery, N. J. The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews* **24**, 581–604 (2000).

## Figures

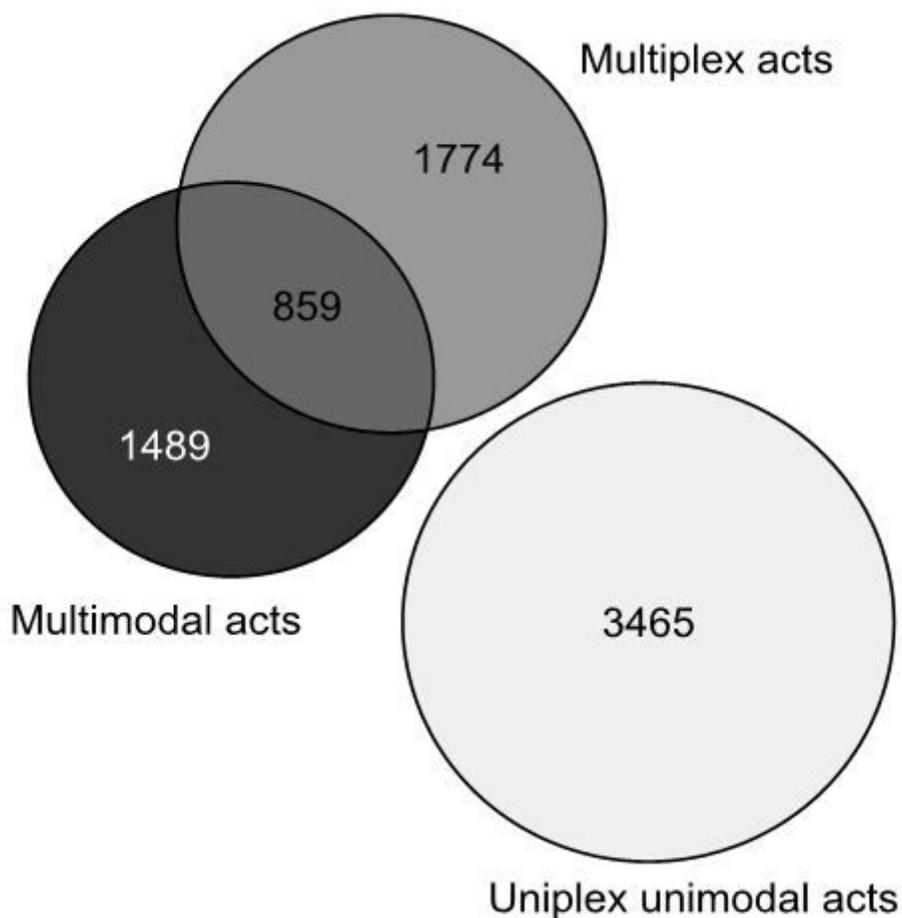
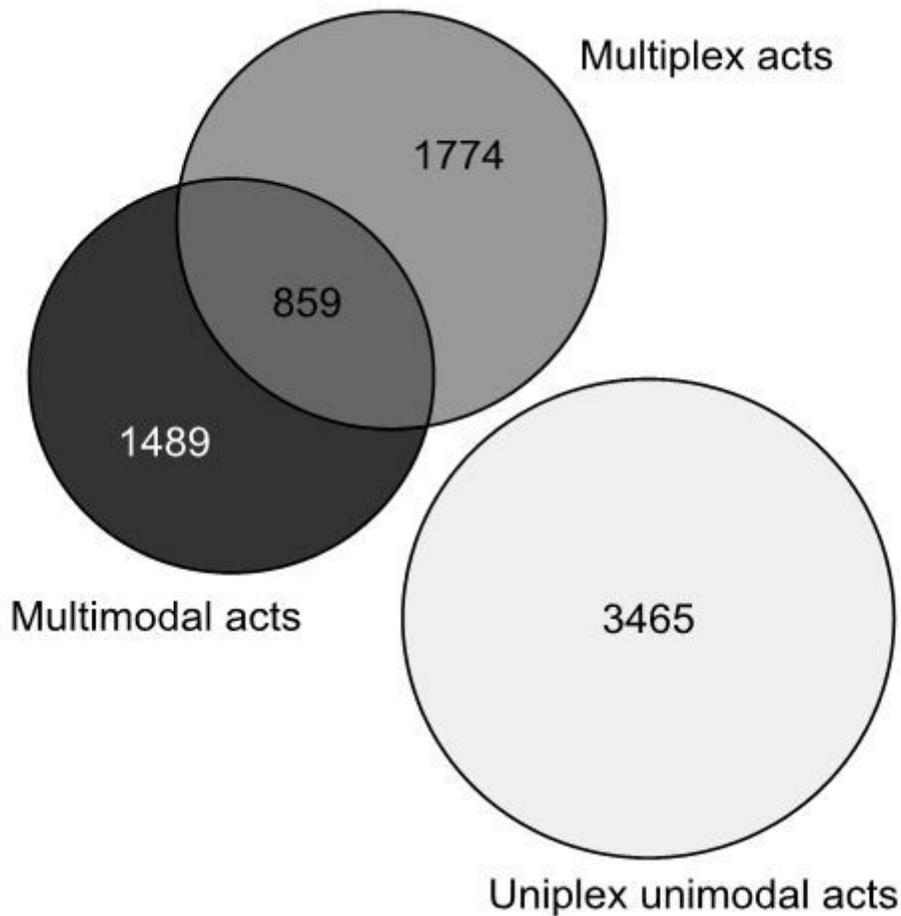


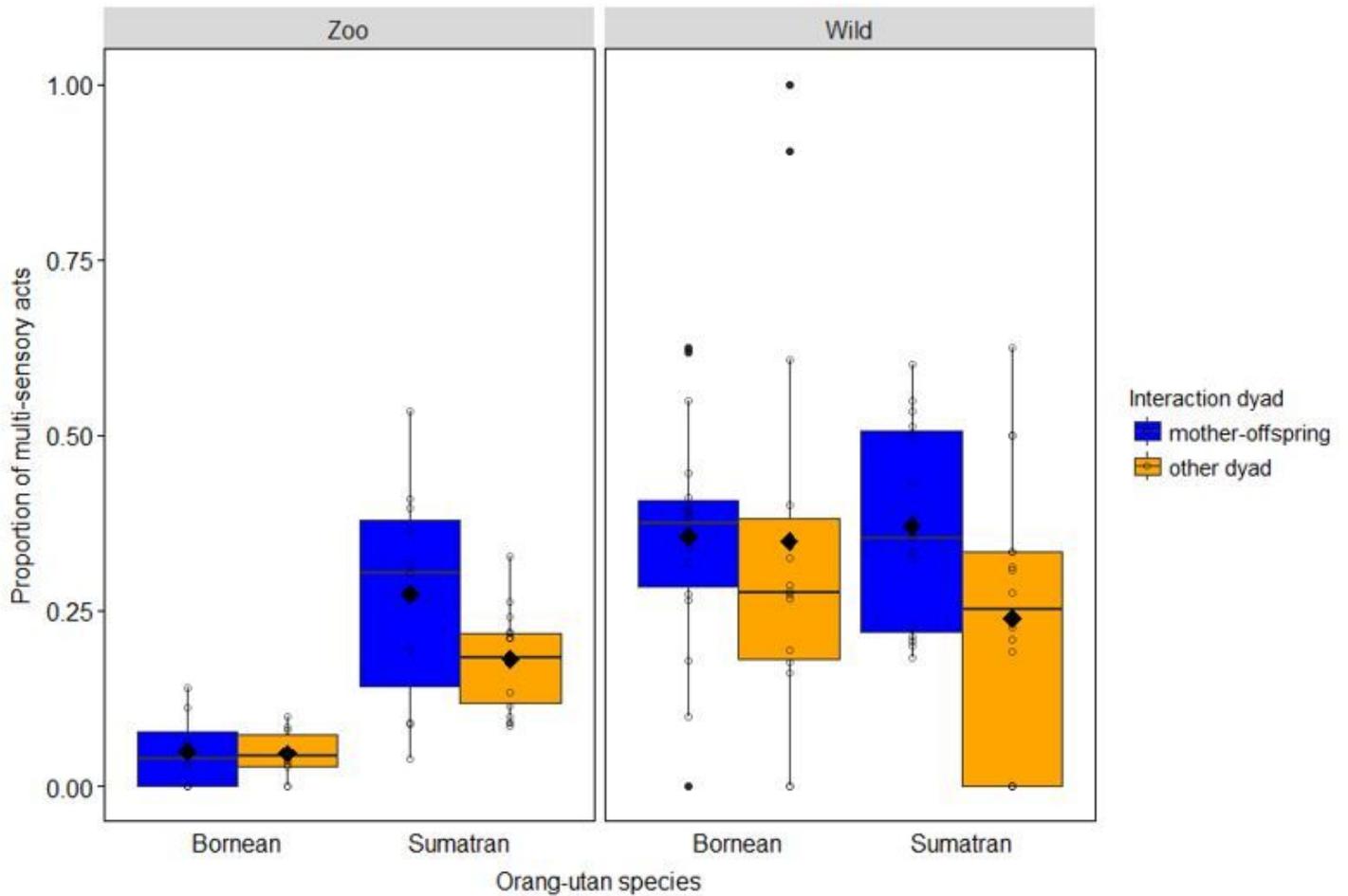
Figure 1

Venn diagram depicting the composition of the dataset with regard to multi-component acts. Depicted values represent sample sizes (N) per type of multi-component communication. Value in overlapping area represents part of the sample that is both multi-sensory (“multimodal”) and multi-articulatory (“multiplex”) (Total N = 7587).



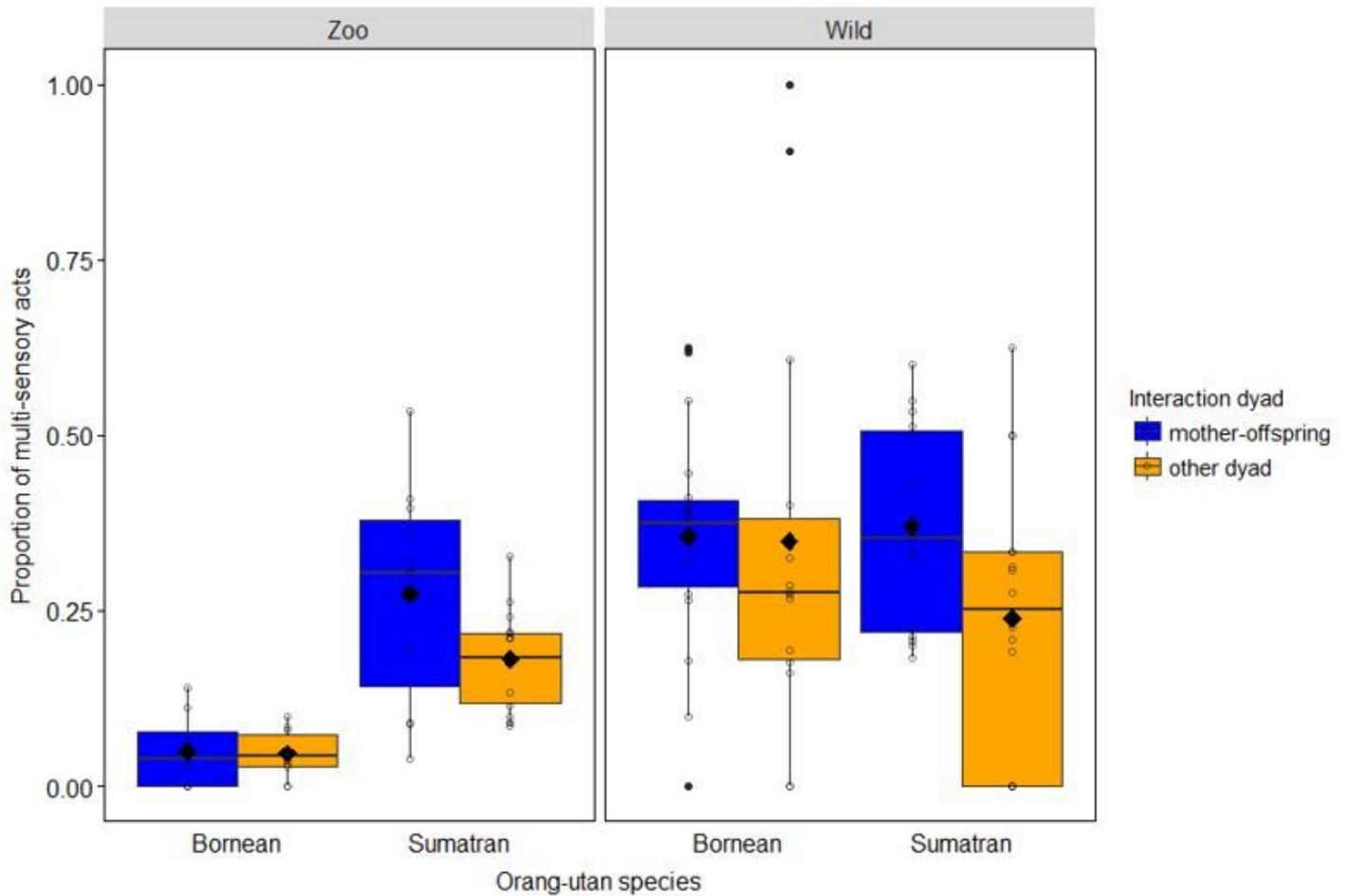
**Figure 1**

Venn diagram depicting the composition of the dataset with regard to multi-component acts. Depicted values represent sample sizes (N) per type of multi-component communication. Value in overlapping area represents part of the sample that is both multi-sensory (“multimodal”) and multi-articulatory (“multiplex”) (Total N = 7587).



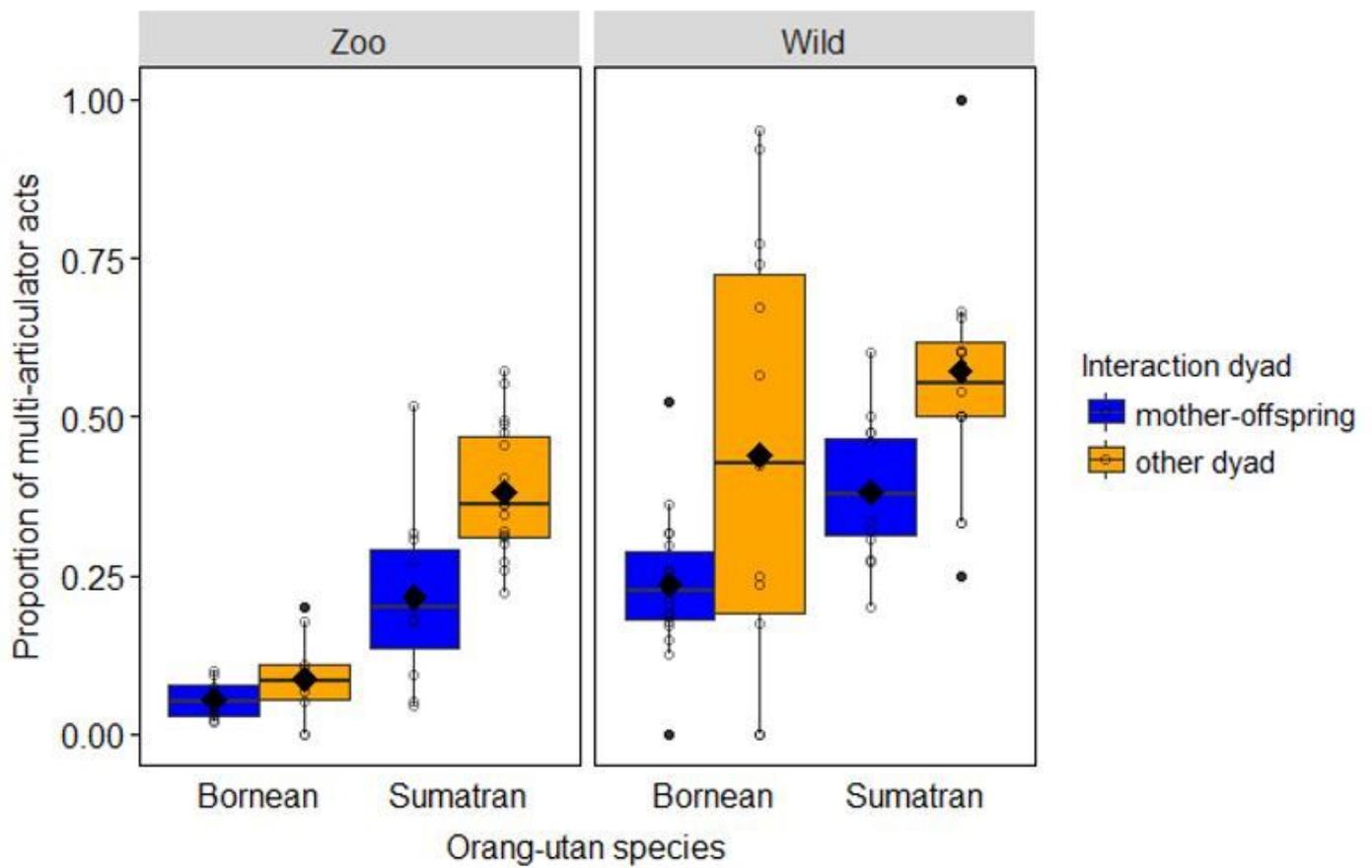
**Figure 2**

Proportion of multi-sensory communicative acts as a function of research setting, species and interaction dyad (“other dyad” includes maternal kin). Indicated are individual means (circles), population means (filled diamonds), medians (horizontal lines), quartiles (boxes), percentiles (2.5% and 97.5%, vertical lines) and outliers (filled dots).



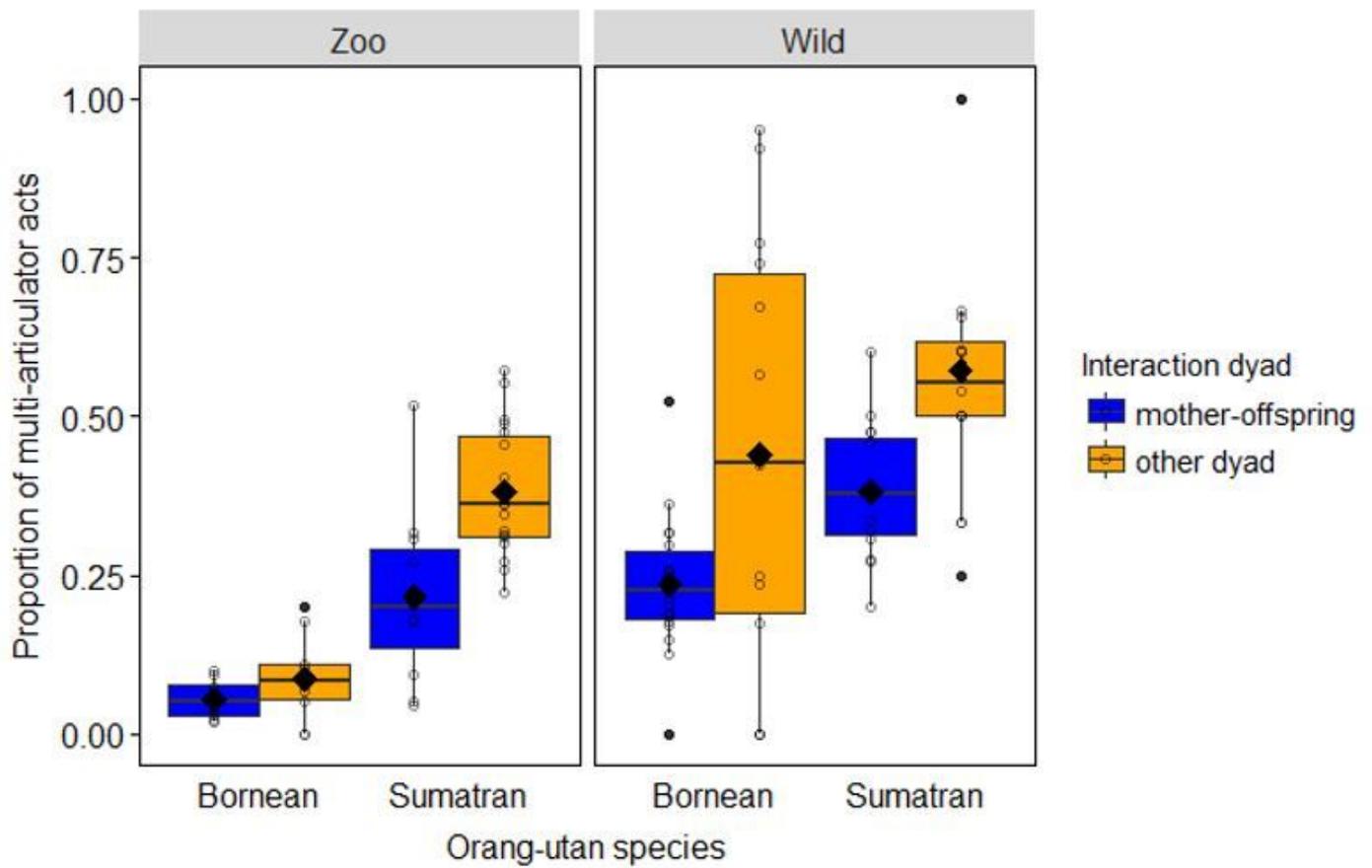
**Figure 2**

Proportion of multi-sensory communicative acts as a function of research setting, species and interaction dyad (“other dyad” includes maternal kin). Indicated are individual means (circles), population means (filled diamonds), medians (horizontal lines), quartiles (boxes), percentiles (2.5% and 97.5%, vertical lines) and outliers (filled dots).



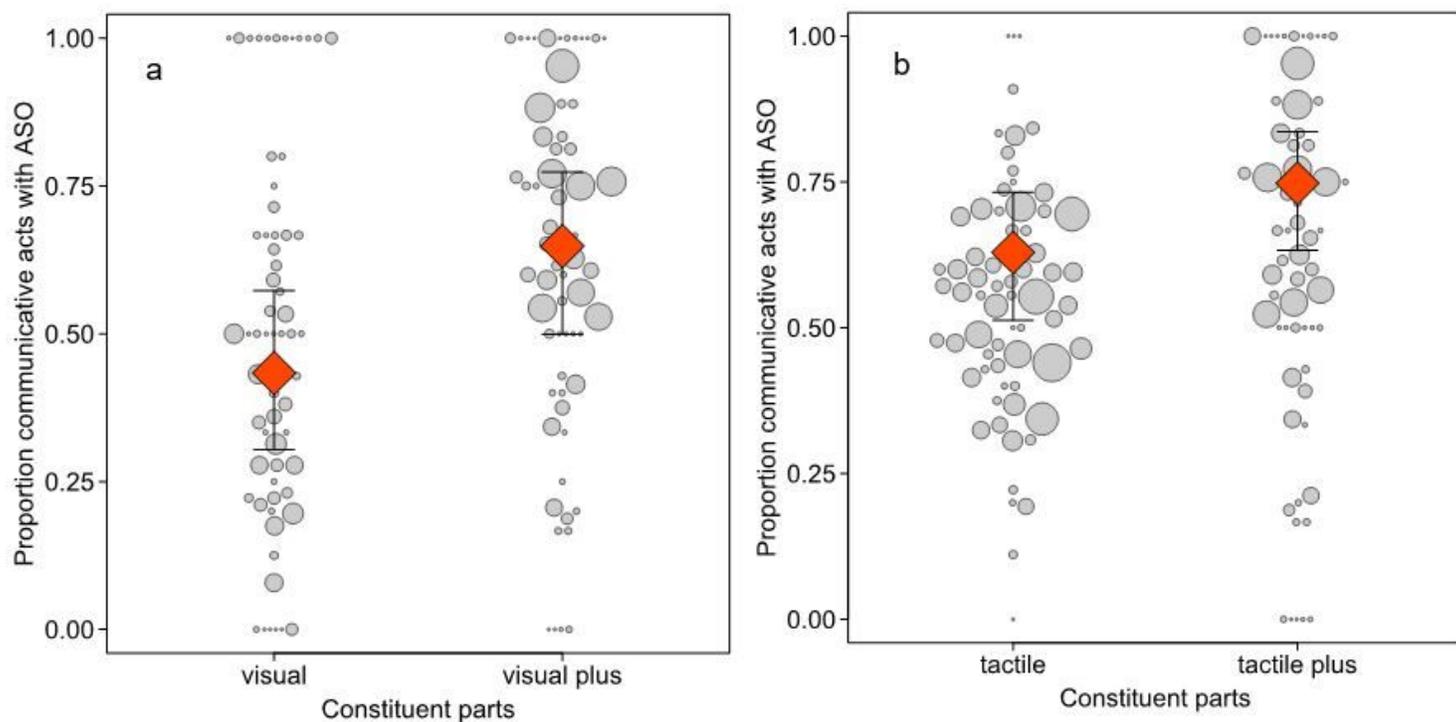
**Figure 3**

Proportion of multi-articulator communicative acts as a function of research setting, species and interaction dyad. Indicated are individual means (circles), population means (filled diamonds), medians (horizontal lines), quartiles (boxes), percentiles (2.5% and 97.5%, vertical lines) and outliers (filled dots).



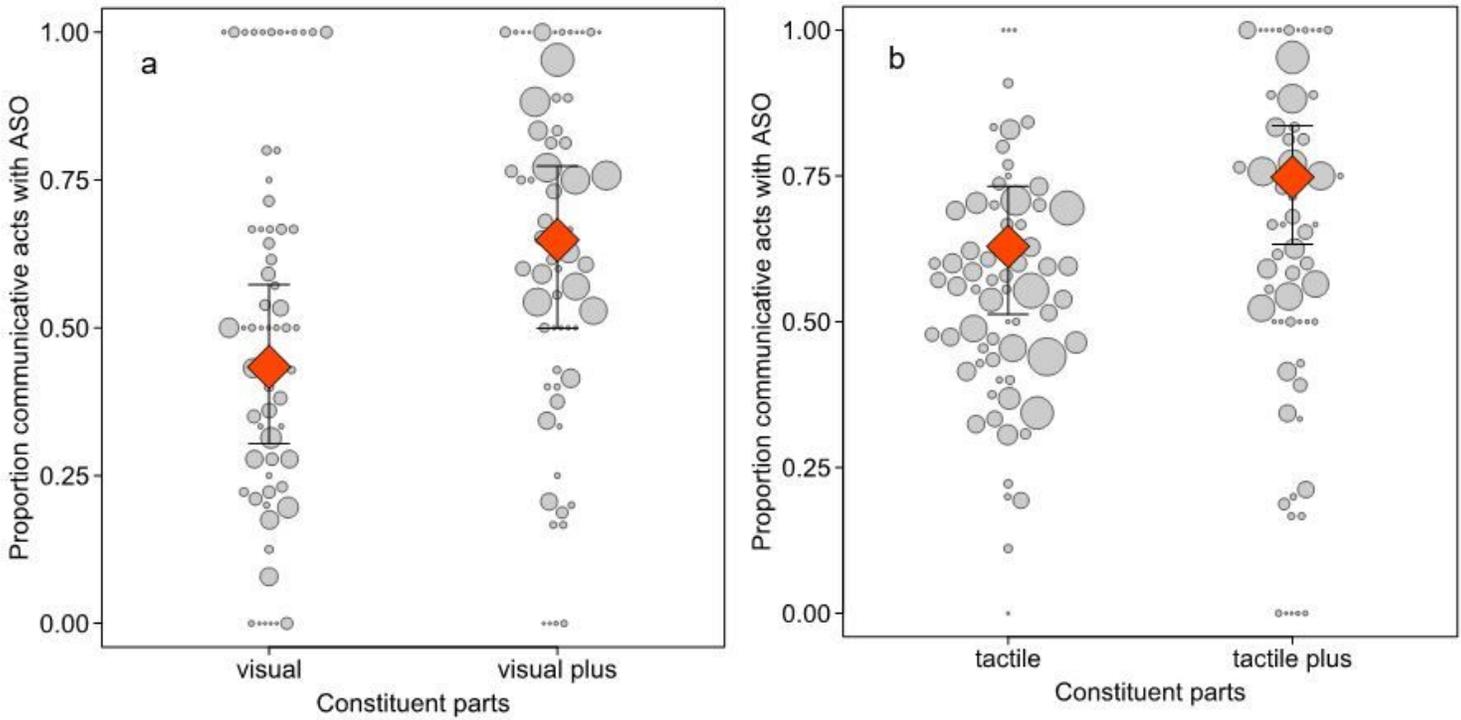
**Figure 3**

Proportion of multi-articulator communicative acts as a function of research setting, species and interaction dyad. Indicated are individual means (circles), population means (filled diamonds), medians (horizontal lines), quartiles (boxes), percentiles (2.5% and 97.5%, vertical lines) and outliers (filled dots).



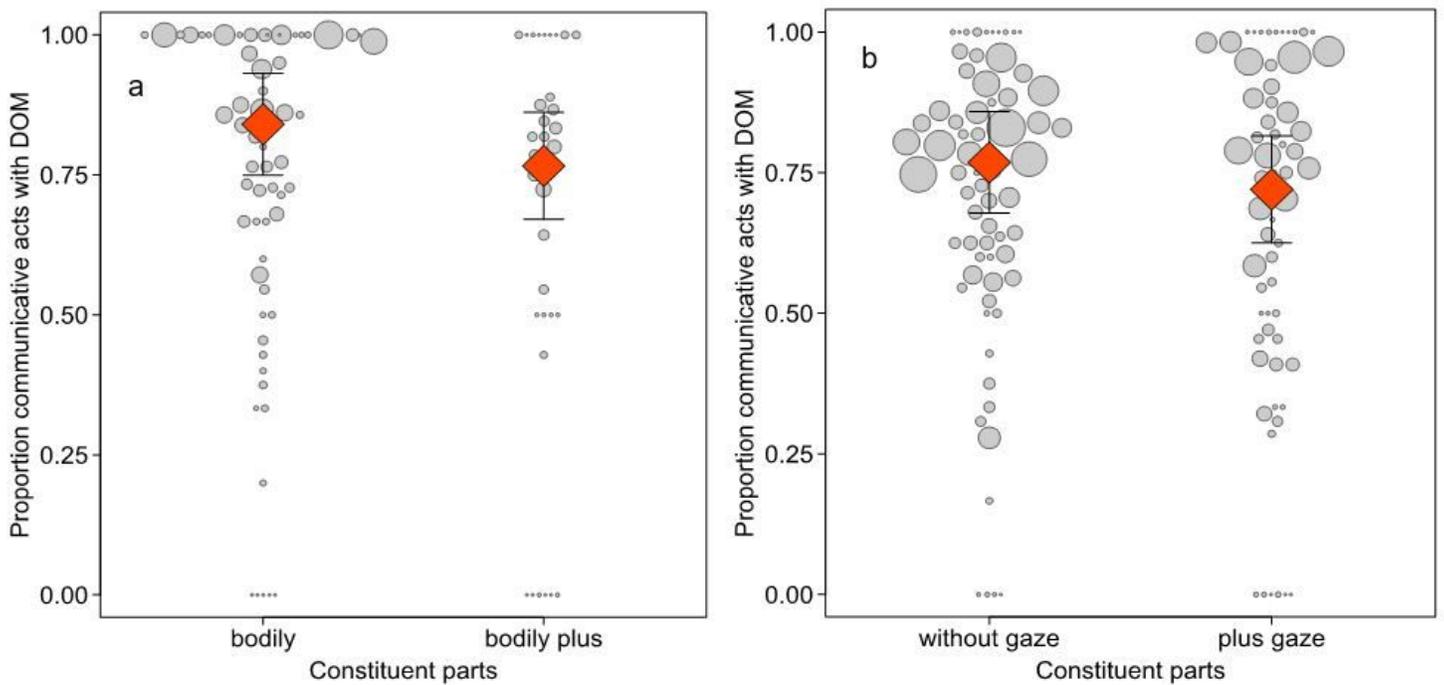
**Figure 4**

Proportion of uni-articulator communicative acts receiving an apparently satisfactory outcome (ASO) as a function of constituent parts (a: visual act only/plus other modality, b: tactile act only/plus other modality). Circles indicate individual means, with circle area representing sample size (a: range = 1–128, b: range = 1–175). Red diamonds depict model estimates with 95% confidence intervals (all other variables centred to a mean of zero).



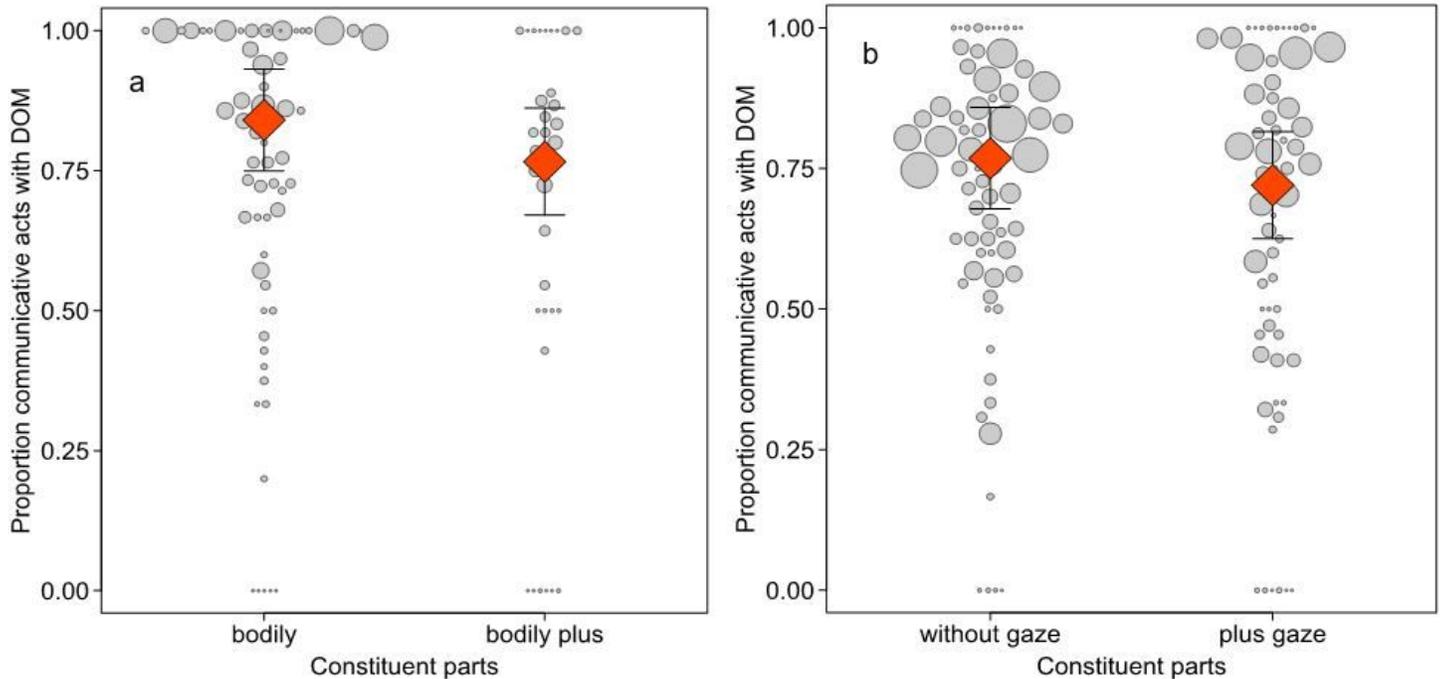
**Figure 4**

Proportion of uni-articulator communicative acts receiving an apparently satisfactory outcome (ASO) as a function of constituent parts (a: visual act only/plus other modality, b: tactile act only/plus other modality). Circles indicate individual means, with circle area representing sample size (a: range = 1–128, b: range = 1–175). Red diamonds depict model estimates with 95% confidence intervals (all other variables centred to a mean of zero).



**Figure 5**

Proportion of communicative acts whose presumed goal matched the dominant outcome (DOM) as a function of different forms of multi-articulator (but uni-sensory) communication (a: bodily act only/plus other articulator, b: CA only/plus gaze). Circles indicate individual means, with circle area representing sample size per individual (a: range = 1–105, b: range = 1–182). Red diamonds depict model estimates with 95% confidence intervals (all other variables centred to a mean of zero).



**Figure 5**

Proportion of communicative acts whose presumed goal matched the dominant outcome (DOM) as a function of different forms of multi-articulator (but uni-sensory) communication (a: bodily act only/plus other articulator, b: CA only/plus gaze). Circles indicate individual means, with circle area representing sample size per individual (a: range = 1–105, b: range = 1–182). Red diamonds depict model estimates with 95% confidence intervals (all other variables centred to a mean of zero).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ESM1TabsFigs.docx](#)
- [ESM1TabsFigs.docx](#)