Article:

Wilke, Claudia, Lahiff, Nicole, Badihi, Gal et al. (9 more authors) (2022) Referential gestures are not ubiquitous in wild chimpanzees: alternative functions for exaggerated loud scratch gestures. ANIMAL BEHAVIOUR. pp. 23-45. ISSN 0003-3472

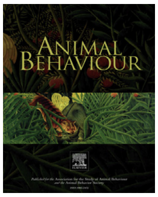
https://doi.org/10.1016/j.anbehav.2022.04.007

Animal Behaviour 189 (2022) 23–45

Contents lists available at ScienceDirect

Animal Behaviour

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



Referential gestures are not ubiquitous in wild chimpanzees: alternative functions for exaggerated loud scratch gestures

C. Wilke [a, 1], N. J. Lahiff [a, 1], G. Badihi [b], E. Donnellan [a, c], C. Hobaiter [b], Z. P. Machanda [d], R. Mundry [e, f, g], S. Pika [h], A. Soldati [c, i], R. W. Wrangham [j], K. Zuberbűhler [b, i], K. E. Slocombe [a, *]

[a] *Department of Psychology, University of York, York, U.K.*
[b] *School of Psychology and Neuroscience, University of St Andrews, St Andrews, U.K.*
[c] *Experimental Psychology, University College London, London, U.K.*
[d] *Departments of Anthropology and Biology, Tufts University, Medford, MA, U.S.A.*
[e] *Cognitive Ethology Laboratory, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany*
[f] *Department for Primate Cognition, Georg-August-University, Göttingen, Germany*
[g] *Leibniz Science Campus Primate Cognition, Göttingen, Germany*
[h] *Institute of Cognitive Science, University of Osnabruck, Osnabruck, Germany*
[i] *Department of Comparative Cognition, University of Neuchâtel, Neuchâtel, Switzerland*
[j] *Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, U.S.A.*

ARTICLE INFO

Article history:
Received 21 December 2021
Initial acceptance 4 February 2022
Final acceptance 2 March 2022
Available online 16 May 2022
MS. number: 21-00713

Keywords:
chimpanzee
communication
cooperation
cultural variation
gesture
grooming
referential signal
scratching

A fundamental aspect of human communication is our ability to refer to external objects and events through both words and gestures (such as pointing), yet the evolutionary origins of such signals remain obscure. Apes, living in their natural environments, rarely or never point, but it has been claimed that male chimpanzees, *Pan troglodytes schweinfurthii*, from the Ngogo community, Uganda, habitually use exaggerated loud scratches (ELSs) to refer to specific body locations where they wish to be groomed (Pika & Mitani, 2006, *Current Biology*, *16*(6), 191–192). This study suggested continuity between referential abilities in humans and our closest living relatives, making it an important finding to replicate in other populations. Hence here, we compared whether ELSs are used in a referential manner across four wild communities of eastern chimpanzees (Ngogo, Kanyawara, Sonso and Waibira). Our data show that scratchers were significantly more likely to receive grooming in the scratched location at Ngogo compared to the other three sites. At the latter sites this response occurred at low rates and signallers did not seem to pursue this goal. This suggests that ELSs do not function referentially at these sites, and the published findings from Ngogo were not replicated. Further exploration into alternative functions of ELSs in the Kanyawara community revealed that, in this community, this signal functions to initiate grooming bouts and to reengage partners during grooming pauses. Individuals who produced the signal to initiate grooming were likely to offer grooming. In contrast, during grooming bouts, groomers produced ELSs to request reciprocation of grooming from their partner. Our study demonstrates that chimpanzees do not ubiquitously use the ELS in a referential manner, but that they can use this gesture in a highly flexible fashion, with signal function depending on the intricate details of the social contexts in which they are produced.

Language remains a complex puzzle in our endeavour to understand human evolutionary history. It is often remarked that language is a hallmark of the human species (Christiansen & Kirby, 2003), and comparative research is fundamental to answer questions such as when and how language evolved, and why it is a uniquely human ability (Hauser et al., 2002). To tackle such questions comparative approaches have often investigated key faculties or 'design features' of language (Hockett, 1960), including semanticity. Referential signals are an important type of semantic signal that direct the attention of recipients to particular aspects of the

---

[*] Corresponding author.
E-mail address: katie.slocombe@york.ac.uk (K. E. Slocombe).

[1] Joint first authors.

https://doi.org/10.1016/j.anbehav.2022.04.007
0003-3472/

environment, where the mapping between signal and meaning is easier to identify than for signals denoting more abstract concepts (Townsend & Manser, 2013; Liebal & Oña, 2018). In children, referential pointing gestures tend to emerge around the same time as first words, and proficiency in early pointing has been considered an important milestone in language development (Bates, 1979; Butterworth, 2003; Kita, 2003; Liszkowski et al., 2012), with many studies suggesting a close link between early pointing and later language abilities (Carpenter et al., 1998; Colonnesi et al., 2010). Once infants start to speak, they can produce word–point combinations (Esteve-Gibert & Prieto, 2014), which tend to precede two-word utterances for external objects (Goldin-Meadow & Butcher, 2003). As speech develops, referential words can then denote a wide range of objects, events, ideas and concepts (Pika, 2012; Sperber & Wilson 1986). Referential communication is therefore a cornerstone of early language acquisition, and it is critical to understand the evolutionary origins of this aspect of language.

Several potential precursors to human referential signals have been identified in nonhuman species, including vocalizations that function as if they refer to external events, such as predator encounters (e.g. Kirchhof & Hammerschmidt, 2006; Seyfarth et al., 1980) and food discovery (e.g. Evans & Evans., 1999; Slocombe & Zuberbühler, 2005). These functionally referential calls have been documented in a wide variety of mammalian and avian species (Gill & Bierema, 2013; Townsend & Manser, 2013). However, it is still unknown whether the cognitive mechanisms underlying the production and reception of these signals are similar to those of humans, since similarity at the behavioural level does not necessarily reflect similarity in the underlying cognitive mechanisms. The potential for differences in psychological mechanisms underpinning signal production and reception has led some scholars to question the importance of functionally referential signals for understanding the evolution of human referential abilities (Wheeler & Fischer, 2012; but see also ; Scarantino & Clay, 2015). In contrast, many nonhuman gestural signals appear to be produced with some degree of intentionality, in that they are goal directed and seem to be produced voluntarily (Liebal et al., 2014). Some gestural signals also importantly have referential qualities, for instance in experimental paradigms where human experimenters provide animals with the opportunity to gesture to request out-of-reach food. Several species have succeeded in such tasks, including domestic horses, *Equus caballus* (Malavasi & Huber, 2016), cetaceans (dolphins, *Tursiops truncatus*: Xitco et al., 2004) and nonhuman apes (hereafter 'apes'; chimpanzees, *Pan troglodytes:* Leavens et al., 2005; Tomasello et al., 1985, 1989; orang-utan, *Pongo pygmaeus*: Call & Tomasello, 1994; bonobos, *Pan paniscus*: Zimmermann et al., 2009). However, accounts of naturally occurring pointing gestures with conspecifics are rarer, particularly in the wild where reports are anecdotal and limited to a handful of cases despite decades of continuous observation of wild populations (bonobos: Vea & Sabater-Pi, 1998; chimpanzees: Hobaiter et al., 2014). Spontaneous referential communication has been claimed for multiple other taxa, including ravens, *Corvus corax*, which 'offer' or 'show' partners objects (Pika & Bugnyar, 2011). Another avian species, Australian magpies, *Gymnorhina tibicen*, have been reported to use a 'pointing' gesture towards model eagles to inform other group members of their presence (Kaplan, 2011). In this case, however, it is difficult to separate the individual simply orienting towards the referent from the richer interpretation that a gesture is produced that provides information to conspecifics. Perhaps surprisingly, some of the most compelling evidence for referential gesturing from a nonprimate subject comes from two coral reef fish species (groupers, *Plectropomus pessuliferus marisrubri*, and coral trout, *Plectropomus leopardus*) that use a 'headstand' signal to indicate hidden prey to potential heterospecific hunting partners (e.g. giant moray eels, *Gymnothorax javanicus*, and octopuses, *Octopus cyanea*; Vail et al., 2013). Although this gesture was seen only occasionally, it produced reliable responses in the recipients and had a clear external referent available. While data on referential gestures in species more distantly related to humans are vital for assessing the selective pressures that made such signalling an adaptive strategy, when trying to trace when in the human lineage referential abilities first emerged, it is important to refocus attention on our closest living relatives, the nonhuman primates.

In nonhuman primates, only a few naturally occurring referential gestures, directed at conspecifics, have been described. For instance, bonobo females produce a 'foot pointing' gesture that appears to refer to their own sexual swelling, resulting in genital–genital rubbing in 83% of cases with the intended recipient (Douglas & Moscovice, 2015). However, it is difficult to ascertain whether the claimed referential function of the signal is understood by recipients, or whether they simply interpret it as a gestural request for action. There may be more similarity between these foot 'points' or leg swing gestures and other widespread 'present genitals' postures females and males perform to solicit copulations than referential points. Another sexual solicitation bonobo gesture, 'beckoning', has been claimed to be spatially referential by indicating a desired location for mating behaviour (Genty & Zuberbühler, 2014). However, in this case the behaviour was rarely produced, and recipients' responses were only in line with the putative referential function in approximately half of the observed cases. In chimpanzees there has been only one study claiming habitual use of a referential gesture. An exaggerated loud scratch (ELS, also termed 'big loud scratch', Hobaiter & Byrne, 2014) is commonly reported across chimpanzee communities (Hobaiter & Byrne, 2011; Goodall, 1986), and is utilized both to initiate joint travel (Hobaiter & Byrne, 2014; also observed in orang-utan: Fröhlich et al., 2019) and in the grooming context (Hobaiter & Byrne, 2014; Pika & Mitani, 2006; also observed in bonobos: Graham et al., 2017). In one community, it was claimed to function as a referential gesture during grooming by indicating a specific body area that the signaller would like their partner to groom and termed the 'directed scratch' (Pika & Mitani, 2006). Presenting body parts to the groomer to fulfil the same function during grooming has been reported in other communities of chimpanzees (Hobaiter & Byrne 2014) and in wild bonnet macaques, *Macaca radiata* (Gupta & Sinha, 2016).

The chimpanzee 'directed scratch' appears to provide some of the most convincing evidence for a customary and frequently used gesture functioning to direct the attention of recipients in a referential manner (Pika & Mitani, 2006). Pika and Mitani (2006) found that in 64% of cases when a groomee produced a directed scratch, they successfully received grooming in the scratched area. However, there are a number of limitations to this study. First, only a small number of the grooming bouts were filmed, and no interobserver reliability was carried out. Second, although Pika and Mitani (2006) assumed the signaller's 'intended meaning' when scratching was to request grooming in a specific location, they did not provide any markers of goal-directed behaviour to confirm this. Rates of persistence or elaboration in the cases where the putative goal was not met are measures that could shed light on this. Finally, their finding was based on adult male–male dyads from a single community of chimpanzees, while investigations of the potentially referential use of this gesture in other dyads (e.g. female–male, female–female), and members of other chimpanzee communities, are absent.

While Pika and Mitani (2006) offered one function for the chimpanzee ELS in a grooming context, it is also important to consider other possible functions for this gesture that have been suggested in the literature. First, Goodall (1986) suggested that exaggerated scratches are used to initiate grooming bouts. This interpretation was supported by Hobaiter and Byrne (2014), who similarly reported that this gesture appeared to be a means to initiate grooming. Hobaiter and Byrne (2014) found that signallers in the Sonso community desisted signalling when grooming started in 82% of cases (indicating an ‘apparently satisfactory outcome’) and scratches appeared to primarily denote ‘groom me’. Second, research across several species suggests that scratching is an indicator of social anxiety. Self-scratching is more common during times of social stress, including after aggressive encounters (pigs, *Sus scrofa*: Norscia et al., 2021; Barbary macaques, *Macaca sylvanus*, and longtailed macaques, *Macaca fascicularis*: Aureli, 1997), during mating seasons when competition is high (ring-tailed lemurs, *Lemur catta*: Sclafani et al., 2012) and when subordinates groom dominant individuals (Barbary macaques: Kaburu et al., 2012). In chimpanzees, anxiety-induced scratching has been discussed for both captive and wild individuals (Baker & Aureli, 1997; Botero et al., 2013). Within the grooming context, which places individuals in vulnerable and physically close positions, it is conceivable that apparent gestural scratches, ELSs, particularly in subordinate individuals, could be conflated with signs of social anxiety.

The aim of this study was to test competing accounts for the function of the ELS during grooming in chimpanzees. Specifically, we examined dyadic interactions to first investigate the ‘referential gesture hypothesis’ that posits that (1a) the ELS functions as a referential gesture to denote the area the signaller wishes to receive grooming in, as reported by Pika and Mitani (2006). For this, we examined data from three communities of eastern chimpanzees (Kanyawara, Sonso and Waibira) alongside the published data from the Ngogo community (Pika & Mitani, 2006). Support for the referential function of scratching would be provided by the signaller receiving grooming within 5 cm of the scratched area, within 10 s of the scratch ending (operational definition applied to the video data from Kanyawara, Sonso and Waibira, which was similar to the original Ngogo data collection; S. Pika, personal observation). To test whether (1b) the goal of the signaller was to elicit grooming in the scratched area, we used established criteria for goal-directed intentional signal production: the signaller should persist or elaborate in their communication, or ultimately terminate the grooming interaction more often when their putative goal is not met (they do not receive grooming in the scratched area) compared to when it is met (grooming received in the scratched area; Townsend et al., 2017). To explore alternative functions for the ELS gesture, we examined the video data from Kanyawara in more detail to test (2a) whether the ELS is used to initiate grooming bouts, as reported by Goodall (1986), Pika (2014) and Hobaiter and Byrne (2014). If this hypothesis is true, we expect a higher rate of ELSs in the period immediately prior to grooming initiation than during the grooming bout. In line with Hobaiter and Byrne's (2014) suggestion that the ELS means ‘groom me’, we hypothesized (2b) that the signaller was more likely to receive grooming rather than give grooming at the beginning of the grooming bout. Once a grooming bout had commenced, we wanted (3) to explore the possibility that ELSs are used to reinitiate grooming when grooming stops, predicting that the rate of ELS will be higher in a grooming pause than the rest of the grooming bout. Following this, we wanted to explore whether (4a) the ELS could function to request grooming (Hobaiter & Byrne, 2014), not only during initiation but also during grooming. We predicted that groomers will be more likely to receive a favourable grooming role switch (receive grooming) in the 10 s following an ELS than during the rest of the grooming bout. Here again, we were able to examine (4b) cases of persistence or elaboration of communication or grooming termination to ascertain whether receiving grooming was the goal of the signaller. Finally, we wanted to investigate the potential that (5) this gesture is a reliable index of signaller anxiety. If the ELS is a sign of social anxiety, we expected individuals to produce more ELSs when grooming with higher-ranking than lower-ranking partners, and with less affiliated than highly affiliated partners.

METHODS

Study Sites and Subjects

Data were collected from four wild chimpanzee communities that range in two forests located ca. 200 km apart in Uganda. The Kanyawara and Ngogo communities are in the Kibale National Park, western Uganda. Kibale is a semideciduous forest with mixed terrain including grasslands, papyrus swamps and secondary forests. While the Kanyawara group's territory is bordered by agriculture and occupied villages (Wrangham et al., 1996), Ngogo is situated approximately 10 km away towards the centre of the forest. Detailed descriptions of the Kibale forest are provided by Chapman and Lambert (2000) and Struhsaker (1997). The other two communities, Sonso and Waibira, are in the Budongo forest in the western Rift Valley in Uganda. The Budongo Forest Reserve is also a semideciduous tropical forest and possesses similar mixed forests to Kibale (Plumptre & Reynolds, 1994).

In the Kibale forest, the Kanyawara community has been systematically studied since 1987 (Wrangham et al., 1992). The group comprised approximately 57 individuals in 2013 (Muller & Wrangham, 2014), the first study period, and approximately 61 individuals in 2018, the final study period. Specifically, data were collected from 22 male and 21 female chimpanzees whose ages ranged between 4 and 50 years at the time of observation. For the Ngogo community, we used the published data set presented by Pika and Mitani (2006). This community has been observed since 1976, with intensive study beginning in 1995 (Mitani & Watts, 1999). At the time of study this was the largest known community, consisting of approximately 145 chimpanzees. Data were collected on 25 adult male individuals, with ages ranging between 19 and 47 years at the time of observation.

In the Budongo forest, the Sonso community has been the subject of continuous study since 1991 (Newton-Fisher, 1997), with an approximate group size of 69 individuals (Hobaiter et al., 2017). The Waibira community has been observed since 2011 and, accounting for those as yet only partially habituated to human presence, this community has between 100 and 120 individuals (Hobaiter et al., 2017). For our cross-site comparison, the data include 18 females and 16 males from Sonso, and 12 females and 28 males from Waibira, aged 17–34 and 22–41 years, respectively.

Equipment

Observational video data were collected with a Panasonic HDC-SD40/60/90 and Panasonic HC-VX980 model camcorder at Kanyawara, with a Sennheiser MKE 400 external microphone attached. At Sonso and Waibira video data were collected with a Panasonic HDC-SD60 or similar model. Published data from Ngogo were obtained from real-time commentary on grooming interactions dictated into a digital recorder, Audioline VR-500, and opportunistic video footage of grooming bouts using a Panasonic NV-GS 250 model camcorder. Kanyawara videos were coded using Noldus Observer XT 14 event logging software for observational data (http://www.noldus.com/animal-behavior-research).

Ethical Note

This study complied with the ASAB/ABS guidelines for the use of animals in research. The study was purely noninvasive, with audio and video recordings taken from a minimum distance of 7 m, in an effort to avoid influencing the natural behaviour of the individuals and groups. The research adhered to the legal requirements of the countries in which it was conducted and followed the recommendations of the 'Animals (Scientific Procedures) Act 1986', as published by the U.K. government, and the principles of 'Ethical Treatment of Non-Human Primates', as stated by the American Society of Primatologists.

Ethical approval for data collection in Kanyawara was granted by the Biology AWERB Ethics Committee (University of York) and the Institutional Animal Care and Use Committee of Harvard University, Tufts University and the University of New Mexico; ethical approval for data collection in Budongo was granted by the Animal Welfare and Ethics Committee of the University of St Andrews. The Ugandan Wildlife Authority and the Ugandan National Council for Science and Technology granted permission to collect data in Uganda.

Definitions

The following definitions were applied to the video data from Kanyawara, Sonso and Waibira communities with a view to being able to generate data sets that were comparable to the published Ngogo data set.

Exaggerated loud scratch (ELS)

Pika and Mitani (2006) defined this signal as 'a relatively loud and exaggerated scratching movement on a part of his (own) body, which could be seen by his grooming partner'. This is broadly in line with Hobaiter and Byrne's (2011) definition of a 'loud exaggerated scratching movement on the signaller's own body' that was accompanied by behaviour that suggested intentional signalling (response waiting, audience checking, persistence) and termed 'big loud scratch'.

In the present study, to reliably operationalize an ELS, we looked for scratches in the video data that were (1) exaggerated in terms of involving the movement of the entire arm, not just the fingers or wrist, (2) were big in terms of being at least 15 cm in length, (3) involved at least two scratches at the same location, (4) were not part of self-grooming, so did not involve the individual inspecting or grooming the scratched location, and finally (5) for the 'referential gesture hypothesis' we only considered scratches that the grooming partner could feasibly see (given that visual knowledge of the scratched location was likely necessary to provide grooming in that location); however, for all other hypotheses we made no restrictions based on the visibility of this multimodal gesture.

Dyadic grooming bout

We defined a dyadic grooming bout as two individuals grooming with one another; this could be unidirectional or bidirectional. Bouts started as soon as one (or both) individual(s) started grooming the other, and there were no other individuals involved in the bout. Bouts were terminated when either individual started grooming a third individual, or when the dyad had ceased grooming for 60 s. If, for instance AT and BO groomed from 01:10 to 02:30 and then from 03:35 to 06:00, then these time periods would be considered two separate grooming bouts.

Initiation period prior to grooming bout

We considered ELSs produced in the 20 s before a grooming bout started as potential initiation signals. While we focus on changes in behaviour within 10 s of ELS signals in the rest of the paper, as individuals could be spatially distant when producing an ELS to initiate grooming, 20 s was chosen for initiation signals to provide sufficient time for individuals to approach following the signal, and to begin grooming.

'Groomee' ELS

This category refers to cases where the signaller was the recipient of grooming before, during and/or after the ELS signal.

Initiation, sequence and response ELS

ELS gestures were video coded as either (1) 'initiation' gestures, meaning neither the signaller nor the partner gave an ELS gesture in the previous 10 s, (2) 'sequence' gestures, meaning the signaller had already given another ELS in the previous 10 s or (3) 'response' gestures, meaning the signaller's partner had already given another ELS in the previous 10 s (sequence took priority over response). Some analyses included only 'initiation' ELSs, whereas others included all ELS types (see Table 1).

Goal-directed behaviours

To test the intended meaning of the ELS from the signaller's perspective we employed methods previously used to identify intentional goal-directed communication (reviewed in Townsend et al., 2017). To test whether an individual had a certain goal when signalling, we looked for an increased likelihood of the signaller persisting or elaborating their communication or terminating the grooming bout if their putative goal was not met, compared to when it was. Persistence was defined as a repetition of the signal within 10 s of the original signal. Elaboration was defined as a different signal with the same goal produced within 10 s of the original signal. Termination of the grooming bout was defined as the signaller terminating the grooming bout within 10 s of the original signal (groomee moves away from the partner; groomer moves away from the partner or stops grooming for at least 60 s). Specific definitions of what constituted the same signal depended on the research question, and these are specified later.

Data Collection

Data were collected at Kanyawara in four study periods (February–May 2013, June 2014–March 2015; July 2016–September 2016; January 2018–October 2018) by C.W. and N.L. To test the 'referential gesture hypothesis', a cross-site analysis was performed, including all four communities. All other hypotheses were addressed with data from the Kanyawara community only, as this is where systematic video data for this project were collected. For this community, in the first two study periods focal adult chimpanzees were video recorded for samples of ca. 15 min, which often included grooming bouts, but not always the initiation and termination of the bouts. In the last two study periods grooming bouts were video recorded with every effort made to capture the whole bout from initiation to termination. Wherever possible, data from both individuals in the dyad were used in the analysis. Overall, 775 dyadic grooming bouts containing 585 ELSs (including those in the 20 s prior to grooming) were collected at Kanyawara, with a total of 2944 min of dyadic grooming from 239 unique dyads.

For the Ngogo community the published data were collected in the period February–March 2005 by S.P. Instances of scratching during male–male grooming bouts were dictated ad libitum on a digital voice recorder, and video footage was collected opportunistically. The groomer ID, groomee ID and any scratches given and their frequency were recorded, as well as the partner's response to the signaller's scratch: whether (1) the partner stopped grooming the spot he was grooming and changed to grooming the scratched location or (2) the partner continued grooming the same location

Table 1
Details of the statistical models constructed to test the research questions

Models and research questions	Response	Test predictors	Control predictors	Random intercepts	Random slopes	*N*	ELS types included
Model 1: referential function							
Does the ELS refer to the location the signaller wishes to receive grooming in, for male-male dyads?	Binomial model Was the target individual (here signaller) groomed within 5 cm of the scratched location: Yes/No	Community*Rank class difference (community has 4 levels: Kanyawara, Ngogo, Sonso, Waibira; rank class difference has 3 levels: lower, same, higher)	Rank class difference, Signaller rank class	Target ID, Partner ID, Dyad ID, Bout ID	None identifiable	279 observations (138 where the partner did groom scratched location); 38 Target IDs, 50 Partner IDs, 89 Dyad IDs, 179 Bout IDs	Initiation, Response
Model 2a: grooming initiation							
Are ELSs given to initiate dyadic grooming?	Poisson model Number of ELSs given by the target individual	Period*Rank difference, Period*CAI (period had 2 levels: during 20 s before grooming start, during the grooming bout; offset by log-transformed period duration)	Rank difference, CAI	Target ID, Partner ID, Dyad ID, Bout ID	CAI, Period, Rank difference, Period*Rank difference and Period*CAI within Target ID and Partner ID. Period and Rank difference within Bout ID and Dyad ID	639 observations; 38 Target IDs, 38 Partner IDs, 102 Dyad IDs, 161 Bout IDs	Initiation, Response, Sequence
Model 2b: grooming initiation; grooming role							
In the 20 s prior to grooming, is the production of an ELS associated with initial grooming role (groomer or groomee)?	Binomial model Grooming role of target individual when grooming commences: Groomer/Groomee	Target individual ELS*Rank difference (Target individual ELS has 2 levels: Yes/No)	Rank difference	Target ID, Partner ID, Dyad ID, Bout ID	Rank difference within Target ID and Partner ID	214 observations (104 where individual became groomer); 33 Targets IDs, 35 Partner IDs, 85 Dyad IDs, 126 Bout IDs	Initiation, Response, Sequence
Model 3: grooming reinitiation when in pause							
Are ELSs given to reinitiate grooming when there is a pause in grooming?	Poisson model Number of ELSs given by target individual	Period*Rank difference Period*CAI (Period has 2 levels: Pause, Active grooming; offset by log-transformed period duration)	Rank difference, CAI	Target ID, Partner ID, Dyad ID, Bout ID	Period, CAI, Rank difference, Period*CAI, Period*Rank difference within Target ID and Partner ID, Period and Rank difference within Dyad ID and Bout ID	1860 observations; 48 Target IDs, 48 Partner IDs, 173 Dyad IDs, 465 Bout IDs	Initiation, Response, Sequence
Model 4a: requesting a favourable grooming switch							
Do groomers give ELSs to elicit grooming from the partner?	Poisson model Number of favourable switches elicited by the target individual (here groomer)	Period*Rank difference, Period*CAI (Period has 2 levels: 10 s after ELS, remainder of duration as groomer in the bout; offset by log-transformed period duration)	Rank difference, CAI	Target ID, Partner ID, Dyad ID, Bout ID	Period, CAI, Rank difference, Period*CAI within Target ID, Period, CAI, Rank difference, Period*CAI, Period*Rank difference within Partner ID	364 observations; 40 Target IDs, 37 Partner IDs, 107 Dyad IDs, 163 Bout IDs	Initiation
Model 4b: requesting a favourable grooming switch; goal-directed behaviour							
Are groomers more likely to persist or elaborate in communication or terminate grooming	Binomial model At least one type of goal-directed behaviour produced by the target individual	Goal met: was there a favourable grooming switch for the target individual within 10 secs of an ELS? (Yes/No)	Rank difference, CAI	Target ID, Partner ID, Dyad ID, Bout ID	Rank difference and CAI within Target ID and Partner ID	299 observations (59 where the goal was met and there was a favourable switch); 40 Target IDs, 37	Initiation

(*continued on next page*)

Table 1 (*continued*)

Models and research questions	Response	Test predictors	Control predictors	Random intercepts	Random slopes	*N*	ELS types included
when the putative goal of eliciting grooming from the partner is not met?	(here groomer) in the 10 s after the initial ELS? (Yes/No)					Partner IDs, 106 Dyad IDs, 162 Bout IDs	
Model 5: social anxiety							
Does an individual produce more ELSs when grooming with a higher-ranking or less affiliated partner?	Poisson model Number of ELSs given by target individual	Rank difference* CAI (offset by log-transformed observation duration)	N/A	Target ID, Partner ID, Dyad ID, Bout ID	Rank difference*CAI within Target ID and Partner ID	1500 observations; 49 Target IDs, 49 Partner IDs, 227 Dyad IDs, 750 Bout IDs	Initiation, Response, Sequence

Full models consisted of test and control predictors along with random intercepts and slopes and were compared to null models which contained only control predictors along with random intercepts and slopes. Models 2–5 were addressed with Kanyawara data only. ELS: exaggerated loud scratch; CAI: composite association index.

he had been grooming before the scratch (for further details see Pika & Mitani, 2006; Pika & Mitani, 2008). Overall, 249 dyadic grooming bouts were observed, containing 186 groomee ELSs from 79 unique male–male dyads.

Data for Sonso and Waibira communities were collected by C.H. between 2011 and 2016 for Sonso, and by C.H., A.S. and G.B. between 2014 and 2020 for Waibira. Here data collection involved focal behaviour sampling (Altmann, 1974) with all social interactions with the potential for gestural communication recorded. While recording was not continuous, a 3 s prerecord function allowed the capture of the onset of any interaction at which the camera was directed. Wherever possible this included recording of the signaller, potential recipients and any subsequent behavioural responses. A running record of the frequency with which particular individuals were observed was maintained and used to select whom to film where multiple targets were available. For Sonso 947 dyadic grooming bouts were identified and checked for groomee ELS, resulting in 68 groomee ELSs from 34 unique dyads; for Waibira 578 grooming bouts were identified and checked, resulting in 45 groomee ELSs from 32 unique dyads.

Video Coding

Kanyawara videos were coded with Observer XT software (version 14.2) to extract continuous details about (1) whether there was dyadic grooming or not, (2) grooming roles, (3) grooming role switches, (4) whether individuals were actively grooming or in a pause, (5) ELS gestures by either individual and (6) ‘present groom’ gestures (‘body is moved to deliberately expose an area to the recipient's attention’; Hobaiter & Byrne, 2014) by either individual.

Testing the Referential Gesture Hypothesis

To investigate this hypothesis, we used data from the four different communities. Once all groomee ELSs had been identified in the videos from Kanyawara, Sonso and Waibira, we manually recorded (1) whether the partner groomed within 5 cm of the scratched area within 10 s after the end of the ELS, (2) whether the signaller produced a ‘present groom’ gesture that overlapped with the ELS, (3) whether the signaller scratched the same area again within 10 s (persistence), (4) whether the signaller gave a ‘present groom’ gesture, presenting the same area as the scratched spot within 10 s (elaboration) and (5) whether the signaller terminated the grooming bout within 10 s. Cases were excluded in which (1) the ELS covered the same spot that was already being groomed, (2) the signaller gave a ‘present groom’ gesture that had some temporal overlap with their ELS and (3) the grooming behaviour in the 10 s following the ELS was not adequately visible.

Intercoder Reliability

N.L. and C.W. performed all the Observer XT video coding on the Kanyawara videos. To assess intercoder reliability N.L. and C.W. both coded the same 40 dyadic grooming bouts, from the total of 775 bouts. In sum, 205 min of dyadic grooming were coded for intercoder reliability of the total 2944 min coded (6.96%). Cohen's kappa was determined for all six grooming behaviours outlined in the ‘video coding’ section above, using the Reliability Analysis function in Observer XT. The kappa value for all six grooming behaviours combined for frequency/sequence (tolerance window = 1 s) was 0.86, and for duration/sequence 0.96, both indicating excellent levels of coder agreement (Fleiss, 1981).

C.W. and K.S. coded all groomee ELS events for the referential gesture hypothesis, using the data sets collected at Kanyawara, Sonso and Waibira. To assess intercoder reliability, C.W. and K.S. both coded the same 20 of the total 230 groomee ELS events (8.70%). Cohen's kappa was calculated for all five behaviours outlined in the section above; the overall kappa value obtained was 0.90, indicating excellent levels of coder agreement (Fleiss, 1981).

Dominance Rank Calculations

The dominance rank difference between the target individual and the partner was included in all statistical models (generally as a control predictor); thus, ranks had to be calculated for all four communities. In the cross-site comparison of the referential function of ELSs, the published Ngogo data set used the directionality of pant-grunts and aggressive interactions between February and July 2005 among the males to rank them into low-, mid- and high-ranking categories (Pika & Mitani, 2008). To provide comparable rank groupings from the other sites, Elo-ratings (Neumann et al., 2011) were first calculated based on the methods of Muller et al. (2021) using long-term data on directionality of pant-grunts for males aged 15 years or older (see Appendix for further details). Subadult males were given an Elo score that was one ordinal rank lower than the lowest-ranking adult male. Next, the ranked males were divided into three equally sized groups (high, mid and low ranking, to match the Ngogo data) for each month. If there was one ‘leftover’ individual (e.g. 10 males divided by three), then an extra individual was put in the low-ranking category. If there were two ‘leftover’ individuals (e.g. 11 divided by three) then one extra individual was put in the low- and one in the mid-ranking category.

In the tests of alternative functions of the ELS, conducted with data from the Kanyawara community only, an average ordinal rank was determined for each individual for each month, including females, with 1 being the highest ranking. Individuals who had not yet entered the adult dominance hierarchy were given a rank of $n+1$, where n=the Elo-rating of the lowest-ranking individual, for any given month. For instance, if in a given month there were 20 adults in the dominance hierarchy, with ranks 1–20, then all nonadults were given a rank of 21.

Composite Association Index (CAI)

A composite association index (CAI) was entered into statistical models investigating alternative functions of ELSs in the Kanyawara community, as a control predictor, with the rationale that the nature of the relationship between grooming partners may affect their communication. The CAI combined three measures: (1) party level association, i.e. the frequency with which two individuals were in the same party; (2) 5 m association, i.e. the frequency with which two individuals were within 5 m of one another; and (3) nearest-neighbour association, i.e. the frequency with which two individuals were seen as nearest neighbours. Scores for each measure were standardized against the community average (all possible dyads among independent individuals, i.e. adults and most subadults), with any value above one being above the community average. The standardized scores for each of these three measures were averaged for each dyad to achieve the CAI for that pair of individuals (based on Gilby & Wrangham, 2008). As the study periods at Kanyawara were spread out over several years, and had varying durations, CAI was determined for every 3–4-month period, for all dyads.

Statistical Models

Model 1: referential function

This model addressed the first research question, whether ELSs refer to the location where the signaller wishes to receive grooming, in male–male dyads. The structure of the model is shown in Table 1, but to summarize, we tested the effect of ‘community’ (fixed effect; four levels: Kanyawara, Ngogo, Sonso, Waibira) on the probability of a target individual's ELS eliciting grooming from the partner within 5 cm of the scratched location (Yes/No), within 10 s of the end of the scratch. The target individual in this case was the individual who produced an ELS while receiving grooming (groomee).

Although the effect of community was the main focus, we also wanted to control for the rank class of the target individual and a potential interaction with rank class difference (between target individual and partner) with community. Previous research at Ngogo suggests that rank class differences may influence whether the scratched area is groomed (Pika & Mitani, 2006). Thus, we included target individual rank class (high, medium, low) and the interaction between community and rank class difference (partner being higher, same or lower rank class than the target individual) in the model (Table 1).

As explained in detail in the Statistical Analysis section below, full models comprising test and control predictors (fixed effects) and random effects were tested against null models comprising only control predictors and random effects (see Table 1 for details of all variables entered).

In addition to this model, we were interested in examining whether the signaller would actively pursue the goal of receiving grooming in the scratched location by showing elevated likelihoods of persisting or elaborating in communication and/or terminating the grooming bout if this putative goal was not met. Doing so was possible for Kanyawara, Sonso and Waibira, where videos were available for all groomee ELSs. We did not run a statistical model on these data as the data were very unbalanced across the behavioural categories, which prevented a stable model being fitted. Instead, we present descriptive statistics in the Results.

Model 2a: grooming initiation

To address the second research question, whether ELSs are given in a grooming initiation context, we investigated the effect of ‘period’ (fixed effect, two levels: 20 s before dyadic grooming, during dyadic grooming) on ELS count, for the target individual (Table 1). To control for varying period duration, we included period duration (log-transformed) as an offset term. We included both individuals in the dyad as target individuals, as long as they had the full 20 s period before dyadic grooming available. For instance, if the dyadic grooming bout involved chimp A and chimp B, and both were visible for the full 20 s before grooming commenced, then we recorded the number of ELSs produced by chimp A as the target individual with chimp B as the partner, as well as the number of ELSs produced by chimp B as the target individual with chimp A as the partner.

Although the effect of ‘period’ was the main focus, we also wanted to rule out interactions with rank difference and CAI. Thus, we included the interaction between period and rank difference, and period and CAI, in the model (Table 1).

Model 2b: grooming initiation; grooming role

A further aim of research question 2 was to test the idea that ELSs at grooming initiation meant ‘groom me’ (Hobaiter & Byrne, 2014). Thus, we tested whether producing an ELS immediately prior to a grooming bout commencing was associated with a higher likelihood of becoming the groomee at grooming initiation.

Although the main focus was the effect of the target individual's ELS (Yes/No) on the probability of the target individual subsequently becoming the groomer or groomee, we also wanted to rule out a potential interaction with Elo rank difference (henceforth rank difference). Thus, we included the interaction between ELS and rank difference as the fixed effect (see Table 1). Again, data were included from both individuals in the dyad (i.e. both were entered as the target individual once), as long as they had the full 20 s period before dyadic grooming available.

Model 3: grooming reinitiation when in pause

To address the third research question; whether ELSs are given to reinitiate grooming when there is a pause in grooming, we tested the effect of ‘period’ (fixed effect, two levels: during pause, during active grooming) on ELS count, for the target individual (Table 1). We included period duration (log-transformed) as an offset term. A pause was defined as a break in active grooming between a given dyad of 3–60 s duration.

Although our main interest was the effect of ‘period’ on ELS count, we also wanted to rule out potential interactions with rank difference and CAI. Thus, we included the interactions between period and rank difference, and period and CAI, as fixed effects (see Table 1). Both individuals from the dyad were included as target individuals in this analysis.

Model 4a: requesting a favourable grooming switch

To address the fourth research question, whether ELSs are given by groomers during a grooming bout to elicit grooming from the partner, we tested the effect of ‘period’ (fixed effect, two levels: 10 s after groomer ELS, remainder of duration as groomer) on ‘favourable switch’ count, for the target individual. We included period duration (log-transformed) as an offset term. A ‘favourable switch’ was a grooming direction switch where the target individual (here the groomer) then received grooming (either becoming the groomee or engaging in mutual grooming).

Although the main focus was the effect of ‘period’ we also wanted to rule out potential interactions with rank difference and CAI. Thus, we included the interactions between period and rank difference, and period and CAI, as fixed effects (see Table 1). Both individuals from the dyad were included as target individuals in this analysis, as long as they produced at least one ELS while in the role of the groomer.

Model 4b: requesting a favourable grooming switch; goal-directed behaviour

A further aim of research question 4 was to test whether the goal of the signaller when producing an ELS was to obtain grooming from their partner. We tested whether the putative goal of a favourable switch was met (fixed effect, two levels: Yes/No) on the probability of the target individual (here the signaller) showing at least one behaviour indicative of pursuing this goal in the 10 s after the initial ELS (persistence or elaboration in communication, or termination of the grooming bout; Yes/No). Here persistence was defined as another ELS in any body area, and elaboration was defined as a ‘present groom’ gesture.

Although the main focus was the effect of the goal being met (or not), we also wanted to control for rank difference and CAI. Thus, we included these three fixed effects in the model (see Table 1). Both individuals from the dyad were included as target individuals in this analysis, as long as they produced at least one ELS while in the role of the groomer.

Model 5: social anxiety

To address the fifth research question, whether individuals produce more ELSs when grooming with a higher-ranking or less affiliated partner, suggesting that ELSs may be a sign of social anxiety, we tested the effects of rank difference and CAI on the number of ELSs produced by the target individual during the dyadic grooming bout. We included observation (grooming) duration (log-transformed) as an offset term. We were also interested in the potential interaction between rank difference and CAI, so included the interaction as an additional predictor variable (see Table 1). Both individuals from the dyad were included as target individuals in this analysis.

Statistical Analysis

Generalized linear mixed models (Baayen, 2008) with binomial error distributions and logit link function or Poisson error distributions with log link function (McCullagh & Nelder 1989) were used for all analyses. Whenever a model comprised an interaction, it included all lower-order terms this encompassed.

Quantitative predictor variables were always scaled by z-transforming to a mean of zero and standard deviation of one, to allow for easier interpretation of the model coefficients (Schielzeth, 2010) and to ease model convergence.

For all Poisson models, we were interested in the count of a particular behaviour during two different periods, which varied considerably in observation effort. Hence, we included the (log-transformed) observation time as an offset term in the models (McCullagh & Nelder, 1989). To control for repeated observations of the same individuals and dyads grooming, target ID, partner ID and dyad ID were included as random intercept effects. Additionally, in the cases where there were multiple observations from the same dyadic grooming bout, bout ID was also included as a random intercepts effect. To avoid underestimation of standard errors and to keep type 1 error rates at 0.05, all theoretically identifiable random slopes were included in each model (Schielzeth & Forstmeier, 2009; Barr et al., 2013; see the Appendix for the random slopes included in each model). Estimates of the correlations between the random intercepts and slopes were also initially included. However, if a ‘singular fit’ message suggested some of the random effects terms to be unidentifiable, as indicated by absolute correlation parameters being essentially one (Matuschek et al., 2017), then these correlations were dropped from the model. In all cases where these were dropped, the correlations did not contribute much to the explanatory value of the model (assessed by the difference in log likelihood; the effects of dropping these are reported in the Appendix). Model stability was estimated by dropping the levels of the random effects one at a time from the data and comparing the estimates derived for the same full model fitted to these subsets with those obtained for the full data set (Nieuwenhuis et al., 2012; see Appendix for all stability values). When there were two or more fixed effects, we assessed collinearity using variance inflation factors (VIF; Field, 2005) based on a standard linear model, which did not include interactions (if any were present in the model), nor random effects. Dispersion was also assessed for all nonbinomial models. None of the models showed concerning values for VIF (>10; Field, 2005) or dispersion; the results of these checks can be found in the Appendix.

To ascertain the effect of the test predictors and to avoid cryptic multiple testing (Forstmeier & Schielzeth, 2011), full–null model comparisons were carried out, with the null model lacking the test predictors (Table 1) in the fixed effects part, but otherwise being identical to the full model in terms of the random and control fixed effects structure. The effect of individual fixed effects was tested by comparing the full model with reduced models lacking them, one at a time (Barr et al., 2013). For the full–null model comparison, as well as the full–reduced model comparisons, a likelihood ratio test was used (Dobson, 2002; R function drop1 with argument ‘test’ set to ‘Chisq’).

Models were implemented in R (version 4.0.2; R Core Team; https://www.R-project.org); using the function glmer of the package lme4 (version 1.1–21; Bates et al., 2015). VIFs were determined using the function vif of the package car (Fox & Weisberg, 2018).

Model stability was assessed using a function written by R.M. Confidence intervals of the model estimates and fitted values were obtained using parametric bootstraps (function bootMer of the package lme4, $N = 1000$ bootstraps).

RESULTS

Here we report the main findings for each of the research questions and models; further model results including estimates, standard errors, confidence intervals, minimum and maximum of model estimates obtained after dropping levels of the random effects one at a time and estimated standard deviations for the contribution of the random effects can be found in Appendix Tables A1–A19. Descriptive statistics on the body areas scratched during ELSs are also presented in the Appendix.

Model 1: Referential Function

The full model fitted the data better than the null model (likelihood ratio test: $\chi^2_9 = 29.249$, $P < 0.001$). However, as the community*rank class difference interaction was found not to be significant, a reduced model was fitted excluding this interaction, but including community, focal rank class and rank class difference as main effects. The reduced model showed a significant effect of community ($\chi^2_3 = 22.624$, $P < 0.001$). The fixed effect ‘community’ was releveled to explore where the significant differences lay between the communities. It was found that at Ngogo the partner was significantly more likely to groom the location scratched by the signaller than at the other three communities (Fig. 1). There were no significant differences between the other three communities.

For Kanyawara, Sonso and Waibira we also had data on dyads including females (female–female and female–male). When including all dyads, the percentage of groomee ELSs to receive grooming within 5 cm of the scratched location within 10 s remained low and were similar for all dyads and male–male dyads (Fig. 2). Further breakdown of all dyads in male–female, female–male and female–female categories for comparison with male–male dyads is available in the Appendix (see Fig. A1).

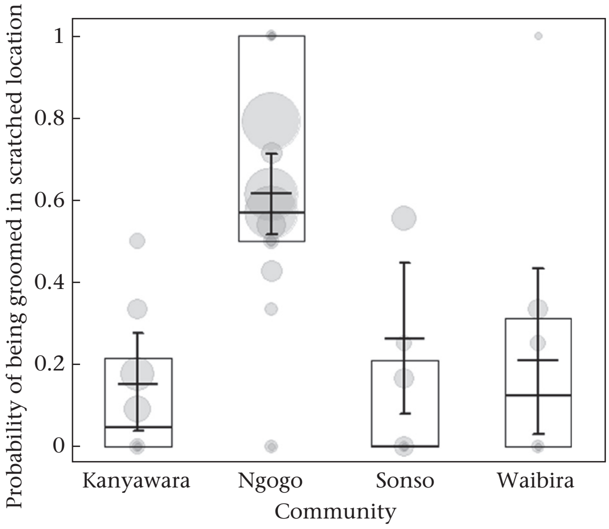


Figure 1. Probability of a groomee ELS eliciting grooming from the partner within 5 cm of the scratched location, for male–male dyads, for each community. Boxes depict median and quartiles for the raw data. Horizontal lines with error bars depict estimates from the fitted model and their 95% confidence limits (when rank class difference and signaller rank class are both manually dummy coded and then centred). The area of the points is proportional to the number of observations for each signaller (range 1–52).

In addition, for Kanyawara, Sonso and Waibira it was possible to descriptively explore whether the signaller's goal (intended meaning) was to elicit grooming in the scratched area. We expected that signallers would be more likely to engage in goal-directed behaviours such as persistence or elaboration of communication or terminate the grooming interaction in cases where the putative goal was not met than when it was met. Instances of these markers of goal-directed behaviour were relatively rare, but in contrast to the pattern expected if the signaller's goal was to elicit grooming in the scratched location, Fig. 3 shows that chimpanzees were more likely in all three communities to show at least one of these three behaviours when they received grooming in the scratched area (putative goal met) compared to when they did not (putative goal not met).

Model 2a: Grooming Initiation

The full model fitted the data better than the null model (likelihood ratio test: $\chi^2_3 = 21.405$, $P < 0.001$). However, as the period*rank difference and period*CAI interactions were found not to be significant, a reduced model was fitted excluding these interactions, but including period, rank difference and CAI as main effects. The reduced model showed a significant effect of period ($\chi^2_1 = 26.834$, $P < 0.001$), indicating that the count of ELSs produced by the target individual was higher in the 20 s period before the start of dyadic grooming compared to the count during the remainder of the grooming bout (see Fig. 4).

Model 2b: Grooming Initiation; Grooming Role

To investigate whether an individual's grooming role when dyadic grooming started was predicted by whether they produced an ELS in the 20 s prior to grooming, we examined this in the cases where an individual was visible for the full 20 s before grooming commenced. We then recorded whether they produced an ELS in this period or not, and whether they became the groomer or groomee when grooming started. Overall, the full model fitted the data better than the null model ($\chi^2_2 = 7.839$, $P = 0.019$). However, as the ELS*rank difference interaction was found not to be significant, a reduced model was fitted excluding this interaction, but including ELS (Yes/No) and rank difference as main effects. The reduced model showed a significant effect of ELS ($\chi^2_1 = 7.558$, $P = 0.006$), as well as rank difference (which was included as a control variable; $\chi^2_1 = 11.685$, $P<0.001$) on subsequent grooming role. Individuals were significantly more likely to become the groomer (rather than groomee) when they gave an ELS in the 20 s before dyadic grooming started (see Fig. 5). When the partner was dominant, individuals were more likely to become the groomer in 65/92 cases (70.65%) than when the partner was subordinate (28/99 cases; 28.28%). In other words, higher-ranking individuals were more likely to be groomed than to groom. This excludes cases where both individuals had the same ordinal rank (e.g. both individuals were subadults).

Model 3: Grooming Reinitiation when in Pause

To explore whether the ELS might also play a role in reinitiating grooming, we analysed the number of ELSs produced in pauses compared to during active grooming. The full model fitted the data better than the null model ($\chi^2_3 = 70.779$, $P < 0.001$). However, as the period*rank difference interaction was found not to be significant, a reduced model was fitted excluding this interaction, but including period, CAI and rank difference as main effects, as well as the period*CAI interaction. The reduced model showed a significant effect of the interaction between period and CAI ($\chi^2_1 = 4.365$,

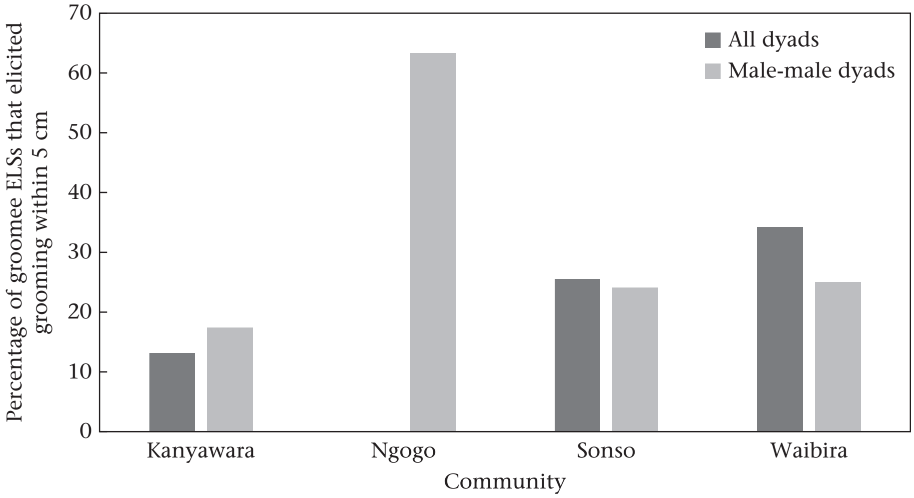


Figure 2. Percentage of groomee ELSs that elicited grooming within 5 cm, within 10 s after the end of the ELS. Results are shown for Kanyawara (*N*=145 groomee ELSs from all dyads, *N*=46 from male–male dyads), Ngogo (no data for all dyads, *N*=188 from male–male dyads), Sonso (*N*=47 from all dyads, *N*=29 from male–male dyads), Waibira (*N*=38 from all dyads, *N*=16 from male–male dyads).

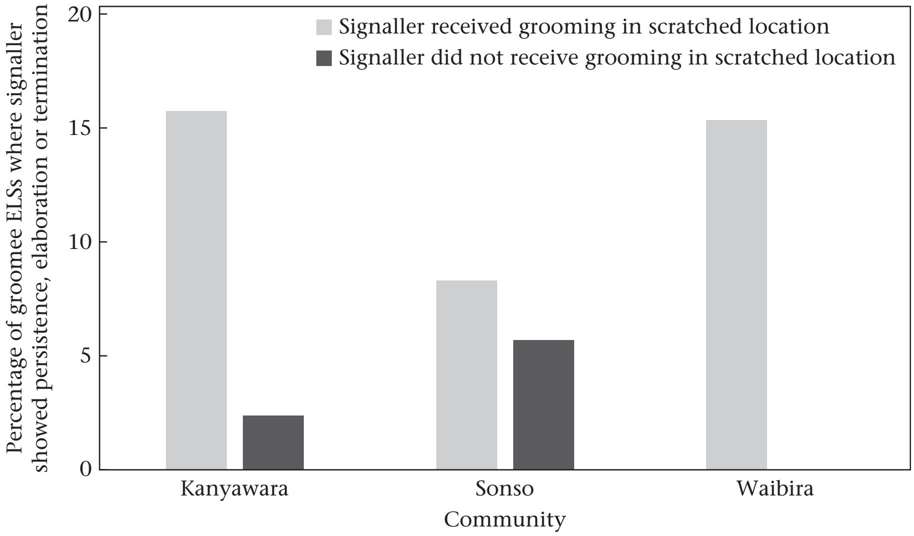


Figure 3. Percentage of groomee ELSs from all dyads where the signaller showed persistence or elaboration of communication and/or termination of the grooming bout in the 10 s following the ELS, as a function of whether the partner groomed the scratched location or not. Results are shown for the three communities where video data were available: Kanyawara (*N*=145 groomee ELSs), Sonso (*N*=47), Waibira (*N*=38).

$P = 0.037$). The count of ELSs was higher in pauses than during active grooming; however, while CAI did not seem to affect ELS count during active grooming, there was a slight increase in ELS count for less affiliated individuals during pauses (Fig. 6).

Model 4a: Requesting a Favourable Grooming Switch

Here we examined whether the groomer might use an ELS to elicit grooming from the partner (i.e. a ‘favourable switch’). The full model fitted the data better than the null model ($\chi^2_3 = 32.877$, $P < 0.001$). However, as the period*rank difference and the period*CAI interactions were found not to be significant, a reduced model was fitted excluding the interactions, but including period, CAI and rank difference as main effects. The reduced model showed a significant effect of period ($\chi^2_1 = 32.390$, $P < 0.001$). The count of favourable switches was significantly higher in the 10 s following an ELS compared to the count during the remainder of the time as a groomer during the grooming bout (see Fig. 7), suggesting that groomers may use this gesture to elicit grooming from their partner during a grooming bout.

Model 4b: Requesting a Favourable Grooming Switch; Goal-directed Behaviour

To ascertain whether the groomer actively pursued the goal of eliciting grooming from the partner (a ‘favourable’ switch) we explored the effect of whether the groomer received grooming in the 10 s after they produced an ELS gesture (Yes/No) on whether they showed at least one goal-directed behaviour, namely persistence or elaboration of communication or termination of the grooming bout. Overall, the full model fitted the data better than the null model ($\chi^2_1 = 15.156$, $P<0.001$). When the signaller produced an ELS but their ‘goal’ of eliciting a favourable grooming switch was not met, they were significantly more likely to show persistence or elaboration of communication and/or termination of grooming (Fig. 8). The control variables rank difference and CAI

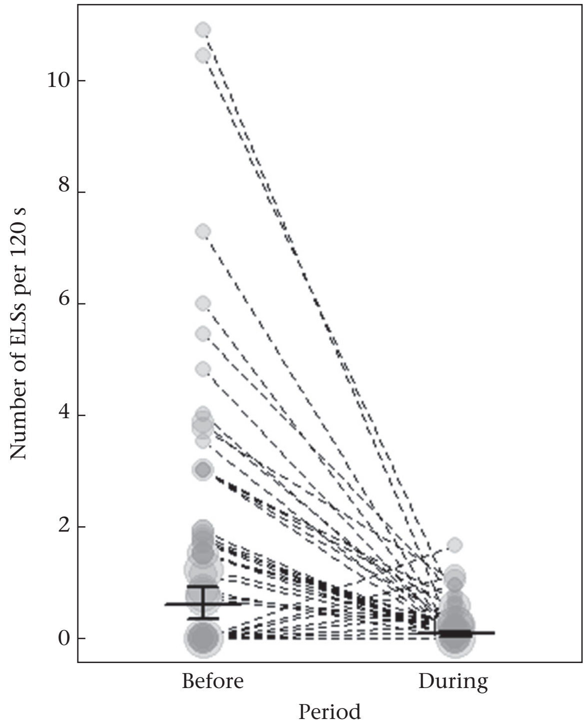


Figure 4. Rate of ELSs given in the 20 s before a dyadic grooming bout, and during the bout, for each dyad. Horizontal lines with error bars depict the fitted model (where number of ELS was offset by log-transformed period duration) and its 95% confidence limits. To effectively visualize the number of ELSs offset by period duration, we illustrate the number of ELSs per 120 s, roughly the average period duration (before and during periods combined). The area of the points is proportional to the number of observations for each dyad for that period (range 1–6).

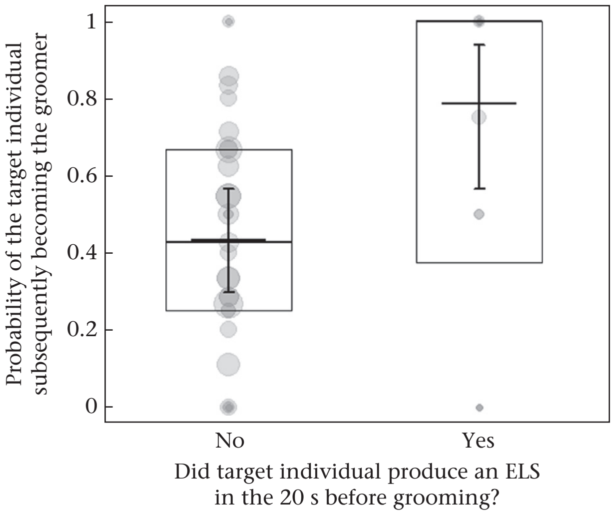


Figure 5. Probability of the target individual becoming the groomer when they did or did not give an ELS in the 20 s before the start of a dyadic grooming bout. Boxes depict median and quartiles for the raw data. Horizontal lines with error bars depict the fitted model and its 95% confidence limits (with average rank difference to the partner). The area of the points is proportional to the number of observations per individual (range 1–15).

were not found to have a significant impact on whether the signaller showed goal-directed behaviours. This model was not very stable in terms of the fixed effect of interest (‘goal met’), meaning it was hard to estimate the effect of this variable conclusively, and the result should be considered with caution. Nevertheless, descriptively, when the goal of receiving a favourable grooming switch was not met following an ELS, the individual then showed goal-directed behaviour in 49/240 cases (20.42%). When a favourable switch was achieved, the individual showed goal-directed behaviour in only 1/59 cases (1.69%).

Model 5: Social Anxiety

Finally, to establish whether ELSs might be a sign of anxiety and more likely to occur in high stress situations, such as grooming with a more dominant and/or less affiliated partner, we explored the effect of rank difference and CAI (and their interaction) with the partner on the number of ELSs an individual produces during grooming. The full model did not fit the data better than the null model ($\chi^2_2 = 2.139$, $P = 0.343$), suggesting that the number of ELSs an individual produced in a dyadic grooming bout was not significantly influenced by the rank difference, the social bond between signaller and partner or the interaction between these. Thus, it seems unlikely that ELSs were merely a sign of social anxiety during dyadic grooming.

DISCUSSION

Our exploration of the function of the ELS signal in grooming contexts failed to replicate the previous finding from Ngogo of this signal functioning in a referential manner in three other chimpanzee communities. We found that signallers in male–male dyads were significantly more likely to receive grooming in the scratched location at Ngogo than in the three remaining sites. In Kanyawara, Sonso and Waibira the likelihood of receiving grooming in the scratched area was too low to consider this a putative function for this signal in these communities, when considering both male–male dyads and all dyads including females. Furthermore, signallers rarely behaved in a manner consistent with pursuing this putative goal. Markers of goal-directed behaviour (persistence or elaboration of communication and termination of the grooming interaction; Townsend et al., 2017) that are expected in the face of not achieving the desired goal of a signal were more frequently observed when the goal of obtaining grooming in the scratched location was met than when it was not met. At Kanyawara, Sonso and Waibira, individuals may rely on another signal, ‘present groom’ (where the signaller's body is moved to deliberately expose an area to the recipient's attention; Hobaiter & Byrne, 2014), to indicate to their partner where they wish to be groomed. Indeed, Hobaiter and Byrne (2014) provided data from Sonso showing that the ‘apparently satisfactory outcome’ (a behaviour change in the recipient that seems to satisfy the signaller and leads to the cessation of signalling) of the ‘present groom’ gesture is to elicit grooming of the presented body part. Similarly, bonnet macaques have also been documented to use present groom gestures for this function (Gupta & Sinha, 2016). Together, these data show that the ELS may only be effective at directing partners to a desired grooming location in the Ngogo community, and therefore that habitual use of a referential gesture in chimpanzees is not ubiquitous in this species. These apparent differences in use have ramifications for our understanding of the evolution of referential signals and highlight the need to continue to look for gestures that may fulfil a referential function in naturalistic conspecific interactions.

There are several reasons why the current exploration of a referential function for the ELS in three different chimpanzee communities may have failed to replicate Pika and Mitani's (2006) original findings. One possibility is that methodological inconsistencies between Pika and Mitani (2006) and the present study led to this discrepancy. Most importantly, scratches from

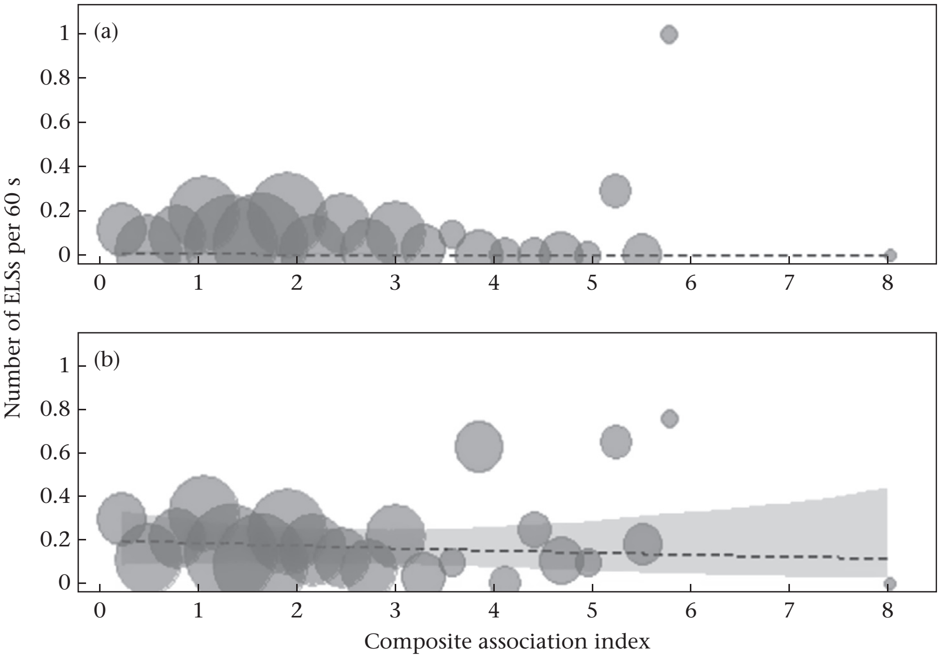


Figure 6. Rate of ELSs produced as a function of the composite association index (CAI) between the dyad partners (a) during active grooming and (b) during a pause. The area of the points is proportional to the number of observations per bin of CAI (range 2–126). The dashed line shows the fitted model and the shaded areas its 95% confidence limits (conditional on rank difference values being at their average). To effectively visualize the number of ELSs offset by period duration, we illustrate the number of ELSs per 60 s, roughly the average period duration (pause and groom periods combined).

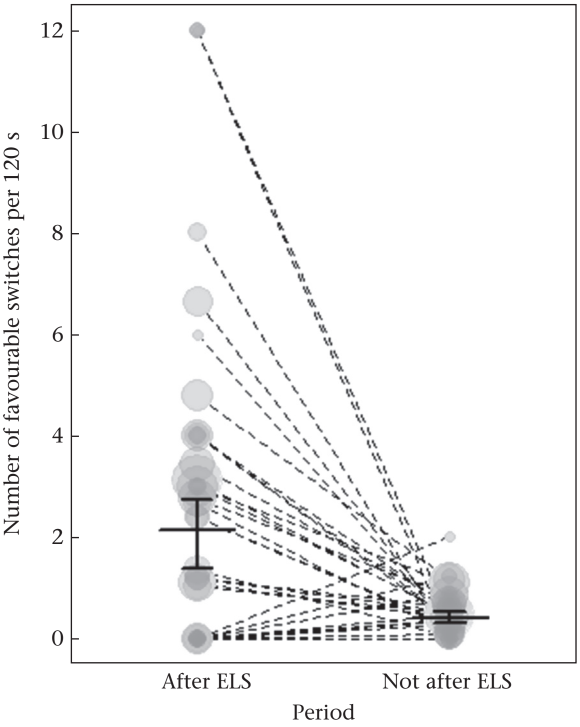


Figure 7. The rate of 'favourable switches' where the groomer received grooming from their partner, in the period immediately following an ELS (after ELS) and during the rest of the dyadic grooming bout (not after ELS). Horizontal lines with error bars depict the fitted model and its 95% confidence limits (when rank difference and the composite association index are at their average). To effectively visualize the number of ELSs offset by period duration, we illustrate the number of ELSs per 120 s, roughly the average period duration (after ELS and not after ELS periods combined). The area of the dots is proportional to the number of observations for each groomer for that period (range 1–20).

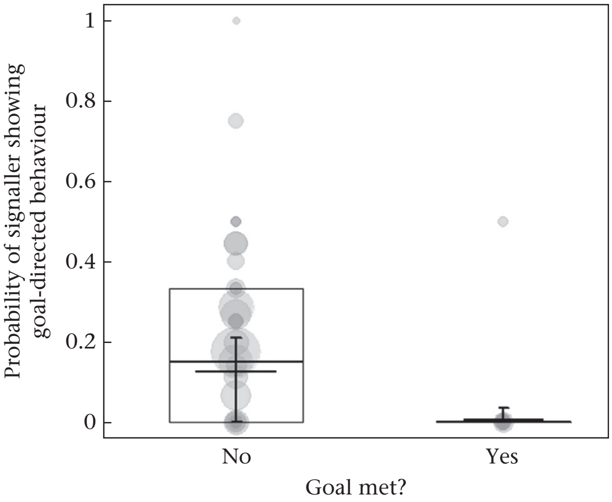


Figure 8. Probability of the signaller showing at least one goal-directed behaviour (persistence or elaboration of communication and/or termination of grooming) when the goal of eliciting a 'favourable' grooming switch was or was not met. Boxes depict median and quartiles. Horizontal lines with error bars depict the fitted model and its 95% confidence limit (at average composite association index and rank difference to the partner). The area of the points is proportional to the number of observations for each signaller (range 1–40).

Kanyawara, Sonso and Waibira were assessed from video data using strict criteria for what constituted an ELS and a 'match' of grooming to that area, whereas the original data set was limited to real-time coding observations from a single observer, alongside a limited amount of video footage that was only examined by one observer. It is crucial that video data coded by multiple observers with robust interobserver reliability measures are collected in the future from the Ngogo chimpanzees to confirm the referential function of this gesture at this site. An additional methodological difference was that sample sizes for Kanyawara, Sonso and Waibira were much smaller for male–male dyads compared to the Ngogo sample, leading to less precise estimates from these communities. However, we also failed to replicate the Ngogo results in the other communities in a larger data set including females in dyads, with low instances of ‘matches’ between scratched and subsequent grooming locations remaining (13% at Kanyawara, 26% at Sonso and 34% at Waibira). Furthermore, given that our descriptive results indicated that individuals from the other communities did not seem to be pursuing the goal of their partner grooming the scratched location, it seems unlikely that more data from these three sites would have altered the pattern observed.

If future work can confirm the referential function of the ELS gesture at Ngogo, this would open the possibility of the ‘directed scratch’ being culturally specific to the Ngogo community. While extensive overlap in the form and broad function of the gestural repertoire of different great ape species has led some to suggest the available repertoires of great ape gesture types are innately specified (Byrne et al., 2017; Graham et al., 2018), it is also the case that they are flexibly deployed to convey different meanings (Hobaiter & Byrne, 2014). Group-specific variation in the presence of attention-getting gestural signals has been claimed in wild chimpanzees (Whiten et al., 1999); however, variation in meaning or function of gestures has not been previously explored in different groups of the same species (Pika & Deschner 2019). In fact, to our knowledge, with the exception of Pika and Mitani's (2006) work on the ELS, data examining the meaning of gestures in wild chimpanzees have been taken solely from the Sonso community (Roberts et al., 2012; 2013; Hobaiter & Byrne, 2014, 2017; Graham et al., 2018). Our data highlight another potential community level difference in use at the level of more fine-grained meaning. ELSs are used to initiate grooming at both Sonso and Kanyawara, but while scratchers in Sonso subsequently receive grooming in the majority of cases (‘groom me’), the majority of scratchers in Kanyawara subsequently provide grooming to their partner (‘let me groom you’). A potentially cultural distinction in gesture meaning in chimpanzees would represent a novel aspect of continuity with human symbolic gestural usage, as diversity in the meaning of our communicative signals is relatively common. For instance, to European citizens a ‘thumbs up’ would be interpreted as an agreement or perhaps ‘good luck’, but in some cultures, such as Iranian culture, this serves as a derogatory gesture (Archer, 1997; Najarzadegan, 2016). While our data present a potentially exciting prospect for the future of ape gestural research, there remain a number of alternative explanations for groupwise variation in the frequency with which a particular meaning might occur or be detected. Replicated video data and evidence of goal-directed behaviour from the signaller's perspective are needed to confirm apparent flexibility in the use of ELS meanings during grooming across populations.

While our data did not support a referential function for the ELS outside of Ngogo, further investigation of alternative functions in the Kanyawara community did reveal an important role for this gesture in aiding cooperative interactions. In line with Goodall (1986) and Hobaiter and Byrne (2014), we found support for the ELS being used to initiate grooming, both prior to a bout beginning and to reinitiate grooming when it had paused. Here, the ELS allows individuals to communicate a desire to start grooming, a cooperative behaviour vital to maintaining social bonds, social tolerance and for repairing relationships following aggression (Dunbar, 1991, 2010; Terry, 1970). Advertising a willingness to cooperate can result in various fitness-enhancing benefits in the animal kingdom. For instance, grouper fish use ‘head shakes’ to engage moray eels in cooperative hunting behaviour, allowing groupers to be almost five times more likely to capture prey than when hunting alone (Bshary et al., 2006). Recruitment calls are an effective signal for cooperation in both predator mobbing (Manser, 2001; Manser et al., 2001) and territory defence contexts (Gersick et al., 2015). Being able to effectively engage conspecifics (or even heterospecifics) in cooperative acts thus has several advantages to an individual, and our findings suggest a pivotal usage of the ELS in instigating and reengaging partners in one such cooperative behaviour.

At Kanyawara we further explored variation in the use of the ELS across the grooming bout. Here, the ELS can facilitate grooming direction switches: groomers produced ELSs to request grooming from their partner. Groomers also demonstrated higher levels of persistence or elaboration of communication or termination of the grooming bout if the goal of receiving grooming from the partner was not met compared to when it was. The ELS signal may therefore play an important role in negotiating the dynamics of a grooming bout, allowing groomers to indicate when reciprocation of this valuable service should be delivered. Interestingly, while a groomer ELS during a grooming bout seems to indicate ‘groom me’, an ELS produced before the initiation of a grooming bout seems to indicate ‘let me groom you’. This indicates potential for flexibility in the ‘meaning’ of the signal within grooming, depending on the specific behavioural context the signal is produced in. Receivers likely have to combine the signal with fine-grained contextual cues to respond in these two differential ways. Future research would need to confirm that when producing an ELS prior to a grooming bout, the signaller's goal was to offer grooming to a partner, but, if confirmed, this would demonstrate the importance of context for correctly interpreting signal meaning (Graham et al., 2020). The integration of context with signals to access the signaller's meaning is fundamental to language (Scott-Phillips, 2015). For instance, stating ‘Michael's on fire!’ would garner very different responses if he had just scored his fifth goal in a football match, compared to if he had just exited a burning building. Using the behavioural context to interpret whether the appropriate response to an ELS signal is to offer grooming or be willing to receive it has the potential to show commonalities with this aspect of pragmatics in language.

While we found promising evidence for two different communicative functions of the ELS in the Kanyawara community (initiate or reinitiate grooming and request reciprocation of grooming), we found no support for these exaggerated scratch gestures being conflated with indications of anxiety. The number of ELSs produced was not influenced by either the rank difference between partners or the social bond between the signaller and partner. Since chimpanzees have been noted to scratch during times of social stress (Baker & Aureli, 1997; Botero et al., 2013), it is likely that exaggerated scratch gestures differ in form from anxiety-induced scratches, or possibly that the particular interactions we focused on were not in fact a source of anxiety for individuals.

The importance of the current work is not limited to identifying the function of the ELS, but also in emphasizing the necessity of both replication and collaboration during data collection. Research on nonhuman primates is prone to multiple methodological criticisms, such as small sample sizes, pseudoreplication and difficulties in directly comparing findings across studies (Slocombe et al., 2011; Waller et al., 2013). The importance of validating previous findings is therefore particularly paramount in a literature vulnerable to false positive results. Often, replication has been seen as devoid of innovation in scientific fields (Nosek et al., 2012), but here we show that replication can reveal interesting new avenues of research. Comparing signal use, function and meaning across populations may reveal important insights into the flexibility of communication systems; however, collaboration between research groups will be vital to ensuring such cross-site comparisons are feasible.

In conclusion, we found significant cross-site variation in the extent to which recipients of groomee ELSs respond by grooming the scratched area. We found little evidence that this signal fulfils a referential function in Kanyawara, Sonso or Waibira communities from either a signaller or a recipient point of view. This finding indicates that habitual use of a referential scratch signal in conspecific grooming interactions is not ubiquitous in chimpanzees, which has implications for our understanding of the evolution of referential signals. In the Kanyawara community the ELS fulfils two alternative functions: first, to initiate grooming bouts and to reinitiate grooming during pauses, and second, for the groomer to request reciprocation of grooming from their partner. The ELS plays an important role in this community of chimpanzees in initiating and negotiating the dynamics of these crucial cooperative dyadic interactions.

Author Contributions

C. Wilke: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing—Original Draft, Writing—Review & Editing, Visualization. **N.J. Lahiff**: Conceptualization, Methodology, Investigation, Data curation, Writing—Original Draft, Writing—Review & Editing. **G. Badihi**: Investigation, Writing—Review & Editing. **E. Donnellan**: Formal analysis, Investigation, Writing—Review & Editing. **C. Hobaiter**: Methodology, Investigation, Writing—Review & Editing. **Z.P. Machanda**: Investigation, Writing—Review & Editing. **R. Mundry**: Formal analysis, Resources, Writing—Review & Editing, Visualization. **S. Pika**: Methodology, Resources, Writing—Review & Editing. **A. Soldati**: Investigation, Writing—Review & Editing. **R.W. Wrangham**: Writing—Review & Editing. **K. Zuberbűhler**: Writing—Review & Editing. **K.E. Slocombe**: Conceptualization, Methodology, Writing—Original Draft, Writing—Review & Editing, Supervision, Project administration, Funding acquisition.

Declaration of Interest

The authors declare no conflicting interests.

Acknowledgments

We are grateful to the field assistants and directors of Kibale Chimpanzee Project and the Budongo Conservation Field Station for assistance with data collection in Uganda. We thank the Ugandan National Council for Science and Technology, the Uganda Wildlife Authority and the President's office for permission to collect data in Uganda and the Royal Zoological Society of Scotland for providing core funding for Budongo Conservation Field Station. We thank Martin N. Muller and Drew K. Enigk for providing Kanyawara dominance ranks, and Abbie Naylor and Grace Sanders for their assistance with video screening and coding. We are grateful to Ruth Sonnweber for statistical discussions. This work was supported by the ERC grant awarded to KS (ERC_CoG 2016_724608).

References

Altmann, J. (1974). Observational study of behaviour: sampling methods. *Behaviour, 49*(3), 227–267. https://doi.org/10.1163/156853974X00534

Archer, D. (1997). Unspoken diversity: Cultural differences in gestures. *Qualitative Sociology, 20*(1), 79–105.

Aureli, F. (1997). Post-conflict anxiety in nonhuman primates: The mediating role of emotion in conflict resolution. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression, 23*(5), 315–328.

Baayen, R. H. (2008). *Analyzing linguistic data*. Cambridge University Press.

Baker, K. C., & Aureli, F. (1997). Behavioural indicators of anxiety: An empirical test in chimpanzees. *Behaviour, 134*(13), 1031–1050.

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

Bates, E. (1979). Intentions, conventions, and symbols. In E. A. Hammel (Ed.), *The emergence of symbols: Cognition and communication in infancy* (pp. 33–68). Academic Press.

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

Botero, M., MacDonald, S. E., & Miller, R. S. (2013). Anxiety-related behavior of orphan chimpanzees (*Pan troglodytes schweinfurthii*) at Gombe National Park, Tanzania. *Primates, 54*(1), 21–26.

Bshary, R., Hohner, A., Ait-el-Djoudi, K., & Fricke, H. (2006). Interspecific communicative and coordinated hunting between groupers and giant moray eels in the Red Sea. *PLoS Biology, 4*(12), e431.

Butterworth, G. (2003). Pointing is the royal road to language for babies. In S. Kita (Ed.), *Pointing* (pp. 17–42). Psychology Press.

Byrne, R. W., Cartmill, E., Genty, E., Graham, K. E., Hobaiter, C., & Tanner, J. (2017). Great ape gestures: Intentional communication with a rich set of innate signals. *Animal Cognition, 20*(4), 755–769.

Call, J., & Tomasello, M. (1994). Production and comprehension of referential pointing by orangutans (*Pongo pygmaeus*). *Journal of Comparative Psychology, 108*(4), 307.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). *Social cognition, joint attention, and communicative competence from 9 to 15 months of age* (pp. i–174). Monographs of the society for research in child development.

Chapman, C. A., & Lambert, J. E. (2000). Habitat alteration and the conservation of african primates: Case study of Kibale national Park, Uganda. *American Journal of Primatology: Official Journal of the American Society of Primatologists, 50*(3), 169–185.

Christiansen, M. H., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in Cognitive Sciences, 7*(7), 300–307. https://doi.org/10.1016/S1364-6613(03)00136-0

Colonnesi, C., Stams, G. J. J., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review, 30*(4), 352–366.

Dobson, A. J. (2002). *An introduction to generalized linear models*. Chapman Hall/CRC.

Douglas, P. H., & Moscovice, L. R. (2015). Pointing and pantomime in wild apes? Female bonobos use referential and iconic gestures to request genito-genital rubbing. *Scientific Reports, 5*, 13999.

Dunbar, R. I. (1991). Functional significance of social grooming in primates. *Folia Primatologica, 57*(3), 121–131.

Dunbar, R. I. (2010). The social role of touch in humans and primates: Behavioural function and neurobiological mechanisms. *Neuroscience & Biobehavioral Reviews, 34*(2), 260–268.

Esteve-Gibert, N., & Prieto, P. (2014). Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Communication, 57*, 301–316.

Evans, C. S., & Evans, L. (1999). Chicken food calls are functionally referential. *Animal Behaviour, 58*(2), 307–319.

Field, A. (2005). *Discovering statistics using SPSS*. Sage.

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

Foerster, S., Franz, M., Murray, C. M., Gilby, I. C., Feldblum, J. T., Walker, K. K., & Pusey, A. E. (2016). Chimpanzee females queue but males compete for social status. *Scientific Reports, 6*(1), 1–11.

Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioural Ecology and Sociobiology, 65*, 47–55.

Fox, J., & Weisberg, S. (2018). *An R companion to applied regression*. Sage.

Fröhlich, M., Lee, K., Setia, T. M., Schuppli, C., & van Schaik, C. P. (2019). The loud scratch: A newly identified gesture of sumatran orangutan mothers in the wild. *Biology Letters, 15*(7), 20190209.

Genty, E., & Zuberbühler, K. (2014). Spatial reference in a bonobo gesture. *Current Biology, 24*(14), 1601–1605.

Gersick, A. S., Cheney, D. L., Schneider, J. M., Seyfarth, R. M., & Holekamp, K. E. (2015). Long-distance communication facilitates cooperation among wild spotted hyaenas, *Crocuta crocuta*. *Animal Behaviour, 103*, 107–116.

Gilby, I. C., & Wrangham, R. W. (2008). Association patterns among wild chimpanzees (*Pan troglodytes schweinfurthii*) reflect sex differences in cooperation. *Behavioral Ecology and Sociobiology, 62*(11), 1831–1842.

Gill, S. A., & Bierema, A. M. K. (2013). On the meaning of alarm calls: A review of functional reference in avian alarm calling. *Ethology, 119*(6), 449–461.

Glickman, M. E., & Doan, T. (2010). *The USCF rating system*. http://www.glicko.net/ratings/rating.system.pdf.

Goldin-Meadow, S., & Butcher, C. (2003). Pointing toward two-word speech in young children. In S. Kita (Ed.), *Pointing* (pp. 93–116). Psychology Press.

Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of behavior*. Belknap Press of Harvard University Press.

Graham, K. E., Furuichi, T., & Byrne, R. W. (2017). The gestural repertoire of the wild bonobo (Pan paniscus): a mutually understood communication system. *Animal Cognition, 20*(2), 171–177.

Graham, K. E., Furuichi, T., & Byrne, R. W. (2020). Context, not sequence order, affects the meaning of bonobo (*Pan paniscus*) gestures. *Gesture, 19*(2–3), 335–364.

Graham, K. E., Hobaiter, C., Ounsley, J., Furuichi, T., & Byrne, R. W. (2018). Bonobo and chimpanzee gestures overlap extensively in meaning. *PLoS Biology, 16*(2), e2004825.

Gupta, S., & Sinha, A. (2016). Not here, there! Possible referential gesturing during allogrooming by wild bonnet macaques, *Macaca radiata*. *Animal Cognition, 19*(6), 1243–1248.

Hauser, M. D., Chomsky, N., & Fitch, T. W. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Hobaiter, C., & Byrne, R. W. (2011). The gestural repertoire of the wild chimpanzee. *Animal Cognition, 14*(5), 745–767.

Hobaiter, C., & Byrne, R. W. (2014). The meanings of chimpanzee gestures. *Current Biology, 24*(14), 1596–1600.

Hobaiter, C., & Byrne, R. W. (2017). What is a gesture? A meaning-based approach to defining gestural repertoires. *Neuroscience & Biobehavioral Reviews, 82*, 3–12.

Hobaiter, C., Leavens, D. A., & Byrne, R. W. (2014). Deictic gesturing in wild chimpanzees (*Pan troglodytes*)? Some possible cases. *Journal of Comparative Psychology, 128*(1), 82.

Hobaiter, C., Samuni, L., Mullins, C., Akankwasa, W. J., & Zuberbühler, K. (2017). Variation in hunting behaviour in neighbouring chimpanzee communities in the Budongo forest, Uganda. *PLoS One, 12*(6), e0178065.

Hockett, C. F. (1960). The origin of speech. *Scientific American, 203*(3), 88–97.

Kaburu, S. S., MacLarnon, A., Majolo, B., Qarro, M., & Semple, S. (2012). Dominance rank and self-scratching among wild female Barbary macaques (*Macaca sylvanus*). *African Zoology, 47*(1), 74–79.

Kaplan, G. (2011). Pointing gesture in a bird-merely instrumental or a cognitively complex behavior? *Current Zoology, 57*(4), 453–467.

Kirchhof, J., & Hammerschmidt, K. (2006). Functionally referential alarm calls in tamarins (*Saguinus fuscicollis* and *Saguinus mystax*)–evidence from playback experiments. *Ethology, 112*(4), 346–354.

Kita, S. (Ed.). (2003). *Pointing: Where language, culture, and cognition meet*. Psychology Press.

Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2005). Intentionality as measured in the persistence and elaboration of communication by chimpanzees (*Pan troglodytes*). *Child Development, 76*(1), 291–306.

Liebal, K., & Oña, L. (2018). Different approaches to meaning in primate gestural and vocal communication. *Frontiers in Psychology, 9*(APR), 1–7. https://doi.org/10.3389/fpsyg.2018.00478

Liebal, K., Waller, B. M., Slocombe, K. E., & Burrows, A. M. (2014). *Primate communication: a multimodal approach*. Cambridge University Press.

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & De Vos, C. (2012). A prelinguistic gestural universal of human communication. *Cognitive Science, 36*(4), 698–713.

Malavasi, R., & Huber, L. (2016). Evidence of heterospecific referential communication from domestic horses (*Equus caballus*) to humans. *Animal Cognition, 19*(5), 899–909. https://doi.org/10.1007/s10071-016-0987-0

Manser, M. B. (2001). The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency. *Proceedings of the Royal Society Series B: Biological Sciences, 268*(1483), 2315–2324.

Manser, M. B., Bell, M. B., & Fletcher, L. B. (2001). The information that receivers extract from alarm calls in suricates. *Proceedings of the Royal Society Series B: Biological Sciences, 268*(1484), 2485–2491.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. Chapman & Hall.

Mitani, J. C., & Watts, D. P. (1999). Demographic influences on the hunting behavior of chimpanzees. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists, 109*(4), 439–454.

Muller, M. N., Enigk, D. K., Fox, S. A., Lucore, J., Machanda, Z. P., Wrangham, R. W., & Thompson, M. E. (2021). Aggression, glucocorticoids, and the chronic costs of status competition for wild male chimpanzees. *Hormones and Behavior, 130*, 104965.

Muller, M. N., & Wrangham, R. W. (2014). Mortality rates among Kanyawara chimpanzees. *Journal of Human Evolution, 66*, 107–114.

Najarzadegan, S. (2016). Intercultural communication: A comparison of Iranian and American nonverbal behaviors. *Journal of Applied Linguistics and Language Research, 3*(3), 75–85.

Neumann, C., Duboscq, J., Dubuc, C., Ginting, A., Irwan, A. M., Agil, M., Widdig, A., & Engelhardt, A. (2011). Assessing dominance hierarchies: Validation and advantages of progressive evaluation with elo-rating. *Animal Behaviour, 82*(4), 911–921.

Newton-Fisher, N. E. (1997). *Tactical behaviour and decision making in wild chimpanzees*. Doctoral dissertation, University of Cambridge.

Nieuwenhuis, R., Te Grotenhuis, H. F., & Pelzer, B. J. (2012). Influence. ME: Tools for detecting influential data in mixed effects models. *The R Journal, 4*, 38–47.

Norscia, I., Collarini, E., & Cordoni, G. (2021). Anxiety behavior in pigs (*Sus scrofa*) decreases through affiliation and may anticipate threat. *Frontiers in Veterinary Science, 8*. https://doi.org/10.3389/fvets.2021.630164

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7*(6), 615–631.

Pika, S. (2012). The case of referential gestural signaling: Where next? *Communicative & Integrative Biology, 5*(6), 578–582.

Pika, S. (2014). Chimpanzee grooming gestures and sounds: What might they tell us about how language evolved? In D. Dor, C. Knight, & J. Lewis (Eds.), *The social origins of language: Early society, communication and polymodality* (pp. 129–140). Oxford University Press.

Pika, S., & Bugnyar, T. (2011). The use of referential gestures in ravens (*Corvus corax*) in the wild. *Nature Communications, 2*(1), 560–565. https://doi.org/10.1038/ncomms1567

Pika, S., & Deschner, T. (2019). A new window onto animal culture: The case of chimpanzee gesturing. *Anthropology of Gesture, 18*(2–3), 237–258. https://doi.org/10.1075/gest.19012.pik

Pika, S., & Mitani, J. (2006). Referential gestural communication in wild chimpanzees (*Pan troglodytes*). *Current Biology, 16*(6), 191–192. https://doi.org/10.1016/j.cub.2006.02.037

Pika, S., & Mitani, J. C. (2008). Referential gestures in chimpanzees in the wild: Precursors to symbolic communication? In A. D. M. Smith, K. Smith, & R. F. i Cancho (Eds.), *The evolution of language: Proceedings of the 7th International Conference (Evolang7)* (pp. 478–479). London: World Scientific Publishing.

Plumptre, A. J., & Reynolds, V. (1994). The effect of selective logging on the primate populations in the Budongo Forest Reserve, Uganda. *Journal of Applied Ecology*, 631–641.

Roberts, A. I., Vick, S. J., & Buchanan-Smith, H. M. (2012). Usage and comprehension of manual gestures in wild chimpanzees. *Animal Behaviour, 84*(2), 459–470.

Roberts, A. I., Vick, S. J., & Buchanan-Smith, H. M. (2013). Communicative intentions in wild chimpanzees: Persistence and elaboration in gestural signalling. *Animal Cognition, 16*(2), 187–196.

Scarantino, A., & Clay, Z. (2015). Contextually variable signals can be functionally referential. *Animal Behaviour, 100*(100), e1–e8.

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology, 20*(2), 416–420.

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution, 1*(2), 103–113.

Sclafani, V., Norscia, I., Antonacci, D., & Palagi, E. (2012). Scratching around mating: Factors affecting anxiety in wild Lemur catta. *Primates, 53*(3), 247–254.

Scott-Phillips, T. C. (2015). Meaning in animal and human communication. *Animal Cognition, 18*(3), 801–805.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Animal Behaviour, 28*(4), 1070–1094.

Slocombe, K. E., Waller, B. M., & Liebal, K. (2011). The language void: The need for multimodality in primate communication research. *Animal Behaviour, 81*(5), 919–924.

Slocombe, K. E., & Zuberbühler, K. (2005). Functionally referential communication in a chimpanzee. *Current Biology, 15*(19), 1779–1784.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142). Harvard University Press.

Struhsaker, T. T. (1997). *Ecology of an African rain forest: logging in Kibale and the conflict between conservation and exploitation*. University Press of Florida.

Terry, R. L. (1970). Primate grooming as a tension reduction mechanism. *Journal of Psychology, 76*(1), 129–136.

Tomasello, M., George, B. L., Kruger, A. C., Jeffrey, M., & Evans, A. (1985). The development of gestural communication in young chimpanzees. *Journal of Human Evolution, 14*(2), 175–186.

Tomasello, M., Gust, D., & Frost, G. T. (1989). A longitudinal investigation of gestural communication in young chimpanzees. *Primates, 30*(1), 35–50.

Townsend, S. W., Koski, S. E., Byrne, R. W., Slocombe, K. E., Bickel, B., Boeckle, M., Braga Goncalves, I., Burkart, J. M., Flower, T., Gaunet, F., Glock, H. J., Gruber, T., Jansen, D. A. W. M., Liebal, K., Linke, A., Miklosi, A., Moore, R., van Schaik, C. P., Stoll, S., … Manser, M. B. (2017). Exorcising grice's ghos : An empirical approach to studying intentional communication in animals. *Biological Reviews, 92*, 1427–1433. https://doi.org/10.1111/brv.12289

Townsend, S. W., & Manser, M. B. (2013). Functionally referential communication in mammals: The past, present and the future. *Ethology, 119*(1), 1–11.

Vail, A. L., Manica, A., & Bshary, R. (2013). Referential gestures in fish collaborative hunting. *Nature Communications, 4*, 1765.

Vea, J. J., & Sabater-Pi, J. (1998). Spontaneous pointing behaviour in the wild pygmy chimpanzee (*Pan paniscus*). *Folia Primatologica, 69*, 289–290.

Waller, B. M., Warmelink, L., Liebal, K., Micheletta, J., & Slocombe, K. E. (2013). Pseudoreplication: A widespread problem in primate communication research. *Animal Behaviour, 86*(2), 483–488.

Wheeler, B. C., & Fischer, J. (2012). Functionally referential signals: A promising paradigm whose time has passed. *Evolutionary Anthropology: Issues, News, and Reviews, 21*(5), 195–205.

Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C. E. G., Wrangham, R. W., & Boesch, C. (1999). Cultures in chimpanzees. *Nature, 399*(6737), 682–685.

Wrangham, R. W., Chapman, C. A., Clark-Arcadi, A. P., & Isabirye-Basuta, G. (1996). Social ecology of Kanyawara chimpanzees: Implications for understanding the costs of great ape groups. In W. C. McGrew, L. F. Marchant, & T. Nischida (Eds.), *Great ape societies* (pp. 45–57). Cambridge University Press.

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

Xitco, M. J., Gory, J. D., & Kuczaj, S. A. (2004). Dolphin pointing is linked to the attentional behavior of a receiver. *Animal Cognition, 7*(4), 231–238. https://doi.org/10.1007/s10071-004-0217-z

Zimmermann, F., Zemke, F., Call, J., & Gómez, J. C. (2009). Orangutans (*Pongo pygmaeus*) and bonobos (*Pan paniscus*) point to inform a human about the location of a tool. *Animal Cognition, 12*(2), 347–358.

Appendix

Methods

Additional details on Elo rating calculation for Sonso and Waibira chimpanzee communities

Sonso and Waibira male ranks were also calculated using the Elo-rating method (following the method used by Muller et al., 2021 for Kanyawara) using long-term observational data on pant-grunts (Sonso: May 2009 to December 2016; Waibira: July 2015 to January 2020). Only pant-grunt interactions between males were included, where both individuals were older than 15 years, and where each was involved in nine or more pant-grunt interactions across the observation period (Glickman & Doan, 2010). A few adjustments had to be made to the calculation method for Sonso and Waibira. (1) k constants were estimated separately for the Sonso and Waibira hierarchies using maximum likelihood fitting (Foerster et al., 2016). For different values of k we calculated the overall log likelihood of the observed data based on winning probabilities for each interaction (Sonso, N=759; Waibira: N=906), selecting the optimal k value as the one that produced the lowest overall log likelihood (smaller overall log likelihood indicates that a given k value produced scores for individuals such that across interactions winners had a higher predicted probability of winning). Using this method, we calculated Elo scores using $k = 67.53$ for Sonso males (N=13) and $k = 69.00$ for Waibira males (N=19). (2) Observational data on pant-grunts included a 6-month global burn-in period (prior to the observation period for ELS) where monthly ranks were not calculated (i.e. Sonso: May 2009 to December 2009; Waibira: July 2015 to January 2016).

Additional details on statistical models

Model 1: Referential function. For this binomial model each row constituted a separate ‘groomee’ ELS case, given by the target individual, and we recorded whether the partner then groomed the target individual (the signaller) within 5 cm of the scratch, within 10 s of the end of the scratch.

MaximumVIF was found to be 1.809, suggesting there was no issue in terms of collinearity between community, signaller rank class and rank class difference. No random slopes were identifiable; thus, there were also no correlations between intercepts and slopes.

We compared the full model, comprising community*rank class difference, community, rank class difference and signaller rank class, with a null model lacking the community*rank class difference interaction and also the main effect of community in the fixed effects part. This model was fitted with the optimizer ‘bobyqa’.

Model 2a. Grooming initiation. For this Poisson model, data were included from both individuals in the dyad (i.e. both were entered as the target individual once), as long as they had the full 20 s period before dyadic grooming available. For instance, if the dyadic grooming bout involved chimp A and chimp B, and both were visible for the full 20 s before grooming commenced, then two rows of data were entered; we recorded the number of ELSs produced by chimp A as the target individual, with chimp B as the partner, as well as the number of ELSs produced by chimp B as the target individual, with chimp A as the partner.

Maximum VIF was found to be 1.00, suggesting there was no issue in terms of collinearity between period, CAI and rank difference. The correlations between random intercept and slopes were dropped for Target ID, Dyad ID and Bout ID; dropping these led to a very minor decrease in model fit (loglikelihoods; full model including the correlation parameters = -254.45 (df=60), full model lacking the correlation parameters = -258.23 (df=39)). Dispersion was assessed; a dispersion parameter of 0.37 suggested the response was underdispersed, and thus more conservative.

The full model comprised period, rank difference, CAI, period*rank difference and period*CAI interactions, with a null model lacking period and the period*rank difference and period*CAI interactions. This model was fitted with the optimizer ‘bobyqa’.

Model 2b. Grooming initiation; grooming role. For this binomial model each row constituted one case where the target individual was visible for the full 20 s before they started dyadic grooming. Data were included from both individuals in each dyadic bout (i.e. both were entered as the target individual once), as long as they both had the full 20 s period before dyadic grooming available. For each case we recorded whether they produced an ELS and the grooming role they subsequently assumed when grooming started. Cases were excluded if (1) the bout started with mutual grooming and (2) the target individual gave a ‘present groom’ gesture in the 20 s prior to grooming. Note that because both individuals in the dyad could not become the groomer within the same bout (this was mutually exclusive), this is a limitation the model cannot account for.

Maximum VIF was found to be 1.00, indicating that there was no issue in terms of collinearity between target individual ELS (Yes/No) and rank difference. The correlations between random intercept and slopes were dropped for Partner ID; dropping this led to practically no decrease in model fit (log-likelihoods: full model with Partner ID correlation parameters = -132.218 (df=12), full model lacking the correlation parameters = -132.232 (df=11)).

We compared the full model, comprising target individual ELS (Yes/No), rank difference and the target individual *rank difference interaction, with a null model lacking target individual ELS and the target individual ELS*rank difference interaction in the fixed effects part. This model was fitted with the optimizer ‘bobyqa’.

Model 3. Grooming reinitiation when in pause. For this Poisson model each grooming bout that contained at least one pause contributed four rows: two for the pause period and two for the active grooming period, as both individuals of the dyad were the target individual once. For each row we recorded the number of ELSs produced by the target individual, as well as the observation time. In the count of ELSs during pauses we excluded ELSs where the target individual was the groomer before the pause and scratched within the first 2 s of the pause, as for these it seemed likely that the pause itself was caused by the groomer stopping to scratch.

Maximum VIF was found to be 1.00, indicating that there was no issue in terms of collinearity between period, rank difference and CAI. The correlations between random intercept and slopes were dropped for Partner ID and Dyad ID; dropping these led to a very minor decrease in model fit (loglikelihoods; full model including the correlation parameters = -551.545 (df=60), full model lacking the correlation parameters = -554.070 (df=42)). Dispersion was assessed, and a dispersion parameter of 0.332 suggested that the response was very underdispersed, thus likely making the model more conservative.

The full model comprised period, rank difference, CAI and the period*rank difference and period*CAI interactions; the null model lacked period, period*rank difference and period*CAI in the fixed effects part. This model was fitted with the optimizer ‘bobyqa’.

Model 4a. Requesting a favourable grooming switch. For this Poisson model, only grooming bouts that contained at least one (initiation) groomer ELS were included. For the individual(s) in the dyad that gave the ELS while in a groomer role (the 'target individual'), there were two rows: one for the 10 s period after an ELS and one for the remainder of their time in a groomer role. For both periods we recorded the number of 'favourable switches' (where the signaller switched grooming role from groomer to groomee or mutual), as well as the available observation time. If both individuals in the dyad gave an ELS while in a groomer role, then each contributed two rows as the target individual for that grooming bout.

Maximum VIF was found to be 1.00, indicating that there was no issue in terms of collinearity between period, rank difference and CAI. The correlations between random intercept and slopes were dropped for Partner ID. Dropping these led to a very minor decrease in model fit (loglikelihoods; full model including the correlation parameters = -351.4871 (*df*=44), full model lacking the correlation parameters = -352.975 (*df*=29)). Dispersion was assessed and a dispersion parameter of 0.794 suggested the response was slightly underdispersed, thus likely making the model more conservative.

We compared the full model, comprising period, rank difference, CAI and the period*rank difference and period*CAI interactions, with a null model lacking period, period*rank difference and period*CAI in the fixed effects part. This model was fitted with the optimizer 'bobyqa'.

Model 4b. Requesting a favourable grooming switch; goal-directed behaviour. For this binomial model each row constituted one case where an individual (the 'target individual') produced an (initiation) groomer ELS. In each case we recorded whether the goal of achieving a favourable switch was met (Yes/No), and whether the target individual showed goal-directed behaviours in the 10 s following this (Yes/No).

Maximum VIF was found to be 1.02, indicating that there was no issue in terms of collinearity between goal met (Yes/No), rank difference and CAI. The correlations between random intercept and slopes were excluded from the model for Target ID and Partner ID. Dropping these led to practically no decrease in model fit (loglikelihoods: full model including the correlation parameters = -121.959 (*df*=18), full model lacking the correlation parameters = -122.006 (*df*=12)).

We compared the full model, comprising goal met (Yes/No), rank difference and CAI, with a null model lacking goal met in the fixed effects part. This model was fitted with the optimizer 'bobyqa'.

Model 5. Social anxiety. For this Poisson model, each grooming bout contributed two rows, as both individuals from the dyad were the target individual once. For each we recorded the number of ELSs given by the target individual, as well as the observation time (dyadic bout duration).

Maximum VIF was found to be 1.00, indicating that there was no issue in terms of collinearity between rank difference and CAI. The correlations between random intercept and slopes were excluded from the model for Target ID and Partner ID. Dropping these led to a minor decrease in model fit (loglikelihoods; full model including the correlation parameters = -815.8644 (*df*=26), full model lacking the correlation parameters = -821.4393 (*df*=14)). Dispersion was also assessed, and a dispersion parameter of 0.475 suggested the response was underdispersed, thus likely making the model more conservative.

We compared the full model, comprising rank difference, CAI and the rank difference*CAI interaction, with a null model lacking rank difference*CAI in the fixed effects part. This model was fitted with the optimizer 'bobyqa'.

Additional Results: The Body Area Scratched during ELS

Although not necessary to answer any of our main research questions, we also coded in the Kanyawara videos the type of body area scratched where a scratch fell unambiguously into one of the following body areas: head/face, arm, leg, front torso and back torso. For Kanyawara, 268 ELSs (*N*=48 scratchers) were coded for body part scratched, which included ELSs given in any grooming role and of any type (initiation, response and sequence).

The front torso was scratched most frequently, 85/268 cases (31.7%), followed by the head/face 71/268 (26.5%), arm 54/268 (20.1%), back torso 51/268 (19.0%) and leg 7/268 (2.6%).

Table A1
Model 1 (referential function): full model

Term	Estimate	SE	Lower CI	Upper CI	Minimum	Maximum
Intercept	−3.566	1.052	−10.841	−2.026	−4.618	-2.846
CommunityNgogo[1]	1.800	0.902	0.274	8.108	0.754	2.551
CommunitySonso[1]	2.100	1.068	−0.181	8.990	1.103	3.004
CommunityWaibira[1]	−14.325	2703.760	−18.832	−10.897	−16.346	−13.406
Rank.class.difflower[2]	1.603	1.234	−5.632	8.466	0.772	2.677
Rank.class.diffsame[2]	1.984	1.044	−0.004	8.620	0.581	2.943
Target.rank.classlow[3]	1.878	0.827	0.287	4.887	1.508	2.242
Target.rank.classmiddle[3]	0.395	0.555	−0.719	1.972	0.194	0.719
CommunityNgogo[1]:rank.class.difflower[2]	0.662	1.252	−5.835	8.044	−0.108	1.798
CommunitySonso[1]:rank.class.difflower[2]	−1.991	1.531	−11.640	5.256	−2.906	−0.968
CommunityWaibira[1]:rank.class.difflower[2]	14.724	2703.761	7.666	20.293	−1.154	16.748
CommunityNgogo[1]:rank.class.diffsame[2]	0.435	1.246	−6.170	3.493	−0.292	2.156
CommunitySonso[1]:rank.class.diffsame[2]	−2.248	1.688	−14.135	1.382	−19.010	−0.611
CommunityWaibira[1]:rank.class.diffsame[2]	14.885	2703.760	10.315	19.818	13.966	17.492

Results from the full model investigating the probability of the partner grooming within 5 cm of the location scratched by the target individual (estimates, SEs, confidence intervals, CI, and minimum and maximum of model estimates when dropping levels of the random effects one at a time).

[1] Dummy coded with community=Kanyawara being the reference category.

[2] Dummy coded with rank.class.diff=higher being the reference category.

[3] Dummy coded with target.rank.class=high being the reference category.

Table A2
Model 1 (referential function): reduced model

Term	Estimate	SE	Lower CI	Upper CI	X^2	*df*	*P*
Intercept	−3.339	0.930	−6.190	−1.728			[1]
Communityngogo[2]	2.184	0.524	1.330	3.815	22.624	3	<0.001
Communitysonso[2]	0.680	0.654	−0.648	2.259			
Communitywaibira[2]	0.392	0.763	−1.667	2.076			
Rank.class.difflower[3]	1.469	0.703	0.027	3.597	7.752	2	0.023
Rank.class.diffsame[3]	1.765	0.675	0.476	3.684			
Target.rank.classlow[4]	1.648	0.747	0.158	3.664	5.588	2	0.061
Target.rank.classmiddle[4]	0.350	0.512	−0.734	1.527			

Results of the reduced model (lacking the interaction) investigating the probability of the partner grooming within 5 cm of the location scratched by the target individual (estimates, SE, confidence intervals, CI, and significance tests). Model stability was only evaluated for the full model.

[1] Not included, as has a very limited interpretation.

[2] Dummy coded with community=Kanyawara being the reference category; the indicated test refers to the overall effect of community.

[3] Dummy coded with rank.class.diff=higher being the reference category; the indicated test refers to the overall effect of rank class difference.

[4] Dummy coded with target.rank.class=high being the reference category; the indicated test refers to the overall effect of target rank class.

Table A3
Model 1 (referential function): estimated standard deviations

Grouping factor	Effect	SD
Bout ID	Intercept	<0.001
Dyad ID	Intercept	<0.001
Target ID	Intercept	0.218
Partner ID	Intercept	<0.001

Estimated standard deviations for the contribution of the random effects for the full model investigating the probability of the partner grooming within 5 cm of the location scratched by the target individual.

Table A4
Model 2a (grooming initiation): full model

Term	Estimate	SE	Lower CI	Upper CI	Minimum	Maximum
Intercept	−5.481	0.341	−6.379	−5.091	−5.776	−5.301
Period[1]	−1.852	0.418	−2.656	−1.002	−2.077	−1.627
Rank.diff[2]	−0.364	0.261	−1.044	0.163	−0.516	−0.234
CAI[3]	−0.132	0.271	−0.746	0.332	−0.261	0.027
Period[1]*rank.diff[2]	0.364	0.320	−0.205	1.136	0.187	0.490
Period[1]*CAI[3]	0.496	0.305	−0.026	1.164	0.352	0.644

Results of the full model investigating the number of ELSs given by the target individual in the 20 s before dyadic grooming, compared to the number during dyadic grooming (estimates, SEs, confidence intervals, CI, and minimum and maximum of model estimates when dropping levels of the random effects one at a time).

[1] Dummy coded with period=before being the reference category.

[2] z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were zero and 10.42, respectively.

[3] Composite association index; z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were 1.95 and 1.33, respectively.

Table A5
Model 2a (grooming initiation): reduced model

Term	Estimate	SE	Lower CI	Upper CI	X^2	*df*	*P*
Intercept	−5.325	0.293	−6.049	−4.913			[1]
Period[2]	−2.096	0.373	−3.082	−1.427	26.834	1	<0.001
Rank.diff[3]	−0.146	0.175	−0.475	0.174	0.707	1	0.400
CAI[4]	0.121	0.171	−0.226	0.451	0.494	1	0.482

Results of the reduced model (lacking the interactions) investigating the number of ELSs given by the target individual in the 20 s before dyadic grooming, compared to the number during dyadic grooming (estimates, SEs, confidence intervals, CI, and significance tests). Model stability was only evaluated for the full model.

[1] Not included, as has a very limited interpretation.

[2] Dummy coded with period=before being the reference category.

[3] z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were zero and 10.42, respectively.

[4] Composite association index; z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were 1.95 and 1.33, respectively.

Table A6
Model 2a (grooming initiation): estimated standard deviations

Grouping factor	Effect[1]	SD
Bout ID	Intercept	0.941
Bout ID	Period	0.985
Bout ID	Rank.diff	0.420
Dyad ID	Intercept	0.100
Dyad ID	Period	0.093
Dyad ID	Rank.diff	0.193
Target ID	Intercept	0.481
Target ID	Period	0.057
Target ID	Rank.diff	0.405
Target ID	CAI	0.055
Target ID	Period*rank.diff	0.047
Target ID	Period*CAI	0.055
Partner ID	Intercept	0.412
Partner ID	Period	0.571
Partner ID	Rank.diff	0.823
Partner ID	CAI	0.590
Partner ID	Period*rank.diff	1.005
Partner ID	Period*CAI	0.426

Estimated standard deviations for the contribution of the random effects for the full model exploring number of ELSs given by the target individual in the 20 s before and during dyadic grooming. CAI: composite association index.

[1] ‘Intercept’ denotes a random intercepts effect; other entries indicate a random slopes effect.

Table A9
Model 2b (grooming initiation; grooming role): estimated standard deviations

Grouping factor	Effect[1]	SD
Bout ID	Intercept	<0.001
Dyad ID	Intercept	<0.001
Target ID	Intercept	0.614
Target ID	Rank.diff	0.452
Partner ID	Intercept	0.649
Partner ID	Rank.diff	<0.001

Estimated standard deviations for the contribution of the random effects for the full model investigating the probability of the target individual becoming the groomer when dyadic grooming started, with this being predicted by whether the target individual produced an ELS before grooming started.

[1] ‘Intercept’ denotes a random intercepts effect; other entries indicate a random slopes effect.

Table A7
Model 2b (grooming initiation; grooming role): full model

Term	Estimate	SE	Lower CI	Upper CI	Minimum	Maximum
Intercept	−0.248	0.256	−0.770	0.242	−0.363	−0.163
Target ELS[1]	1.638	0.651	0.576	4.320	1.336	2.226
Rank.diff[2]	0.788	0.260	0.337	1.413	0.717	1.104
Target ELS[1]* rank.diff[2]	0.331	0.634	−0.928	2.683	0.034	0.804

Results of the full model investigating the probability of the target individual becoming the groomer when dyadic grooming started, with this being predicted by whether the target individual produced an ELS before grooming started (estimates, SEs, confidence intervals, CI, and minimum and maximum of model estimates when dropping levels of the random effects one at a time).

[1] Dummy coded with target individual ELS=No being the reference category.

[2] z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were -0.24 and 9.84, respectively.

Table A8
Model 2b (grooming initiation; grooming role): reduced model

Term	Estimate	SE	Lower CI	Upper CI	X^2	*df*	*P*
Intercept	−0.265	0.260	−0.836	0.221			[1]
Target ELS[2]	1.564	0.615	0.487	3.172	7.632	1	0.006
Rank.diff[3]	0.833	0.255	0.396	1.544	11.525	1	<0.001

Results of the reduced model (lacking the interaction) investigating the probability of the target individual becoming the groomer when dyadic grooming started, with this being predicted by whether the target individual produced an ELS before grooming started (estimates, SEs, confidence intervals, CI, and significance tests). Model stability was only evaluated for the full model.

[1] Not included, as has a very limited interpretation.

[2] Dummy coded with target individual ELS=No being the reference category.

[3] z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were -0.24 and 9.84, respectively.

Table A10
Model 3 (grooming reinitiation when in pause): full model

Term	Estimate	SE	Lower CI	Upper CI	Minimum	Maximum
Intercept	−9.922	0.488	−13.128	−9.322	−10.516	−9.746
Period[1]	4.091	0.486	3.400	7.026	3.805	4.428
Rank.diff[2]	−0.325	0.287	−0.913	0.334	−0.476	−0.203
CAI[3]	−0.800	0.378	−1.645	−0.120	−1.005	−0.660
Period[1]*rank.diff[2]	0.427	0.288	−0.206	1.057	0.298	0.687
Period[1]*CAI[3]	0.690	0.393	−0.040	1.608	0.506	0.909

Results of the full model investigating the number of ELSs given by the target individual during active grooming and during pauses (estimates, SEs, confidence intervals, CI, and minimum and maximum of model estimates when dropping levels of the random effects one at a time).

[1] Dummy coded with period=groom being the reference category.

[2] z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were 8.72 and 6.30, respectively.

[3] Composite association index; z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were 2.06 and 1.39, respectively.

Table A11
Model 3 (grooming reinitiation when in pause): reduced model

Term	Estimate	SE	Lower CI	Upper CI	X^2	*df*	*P*
Intercept	−10.033	0.490	−13.310	−9.397			[1]
Period[2]	4.197	0.485	3.452	7.228			[1]
Rank.diff[3]	0.018	0.166	−0.312	0.381	0.011	1	0.916
CAI[4]	−0.875	0.401	−1.784	−0.233			[1]
Period[2]*CAI[4]	0.783	0.417	0.092	1.781	4.365	1	0.037

Results of the reduced model (lacking the interactions) investigating the number of ELSs given by the target individual during active grooming and during pauses (estimates, SEs, confidence intervals, CI, and significance tests). Model stability was only evaluated for the full model.

[1] Not included, as has a very limited interpretation.

[2] Dummy coded with period=groom being the reference category.

[3] z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were 8.72 and 6.30, respectively.

[4] Composite association index; z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were 2.06 and 1.39, respectively.

Table A12
Model 3 (grooming reinitiation when in pause): estimated standard deviations

Grouping factor	Effect[1]	SD
Bout ID	Intercept	0.958
Bout ID	Rank.diff	0.168
Bout ID	Period	0.394
Dyad ID	Intercept	<0.001
Dyad ID	Rank.diff	<0.001
Dyad ID	Period	<0.001
Target ID	Intercept	1.008
Target ID	Rank.diff	0.227
Target ID	Period	1.080
Target ID	CAI	0.284
Target ID	Period*rank.diff	0.109
Target ID	Period*CAI	0.437
Partner ID	Intercept	0.187
Partner ID	Rank.diff	<0.001
Partner ID	Period	0.100
Partner ID	CAI	<0.001
Partner ID	Period*rank.diff	<0.001
Partner ID	Period*CAI	<0.001

Estimated standard deviations for the contribution of the random effects for the full model investigating the number of ELSs given by the target individual during active grooming and during pauses. CAI: composite association index.

[1] 'Intercept' denotes a random intercepts effect; other entries indicate a random slopes effect.

Table A13
Model 4a (requesting a favourable grooming switch): full model

Term	Estimate	SE	Lower CI	Upper CI	Minimum	Maximum
Intercept	−3.986	0.168	−4.448	−3.760	−4.082	−3.904
Period[1]	−1.732	0.193	−2.018	−1.272	−1.805	−1.610
Rank.diff[2]	−0.221	0.175	−0.579	0.118	−0.298	−0.137
CAI[3]	0.076	0.179	−0.321	0.336	−0.087	0.122
Period[1]*rank.diff[2]	0.122	0.179	−0.197	0.457	0.061	0.202
Period[1]*CAI[3]	0.011	0.178	−0.308	0.411	−0.074	0.147

Results of the full model investigating the number of ‘favourable switches’ in the 10 s after a target individual produces an ELS, compared to the number during the remainder of dyadic grooming (estimates, SEs, confidence intervals, CI, and minimum and maximum of model estimates when dropping levels of the random effects one at a time).

[1] Dummy coded with period=afterELS being the reference category.

[2] z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were 1.40 and 10.35, respectively.

[3] Composite association index; z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were 2.10 and 1.28, respectively.

Table A14
Model 4a (requesting a favourable grooming switch): reduced model

Term	Estimate	SE	Lower CI	Upper CI	X^2	*df*	*P*
Intercept	−3.983	0.168	−4.410	−3.739			[1]
Period[2]	−1.724	0.188	−2.013	−1.305	32.390	1	<0.001
Rank.diff[3]	−0.141	0.131	−0.403	0.116	1.089	1	0.297
CAI[4]	0.082	0.133	−0.190	0.324	0.367	1	0.545

Results of the reduced model (lacking the interactions) investigating the number of ‘favourable switches’ in the 10 s after a target individual produces an ELS, compared to the number during the remainder of dyadic grooming (estimates, SEs, confidence intervals, CI, significance tests, and minimum and maximum of model estimates when dropping levels of the random effects one at a time). Model stability was only evaluated for the full model.

[1] Not included, as has a very limited interpretation.

[2] Dummy coded with period=afterELS being the reference category.

[3] Composite association index; z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were 1.40 and 10.35, respectively.

[4] z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were 2.10 and 1.28, respectively.

Table A15
Model 4a (requesting a favourable grooming switch): estimated standard deviations

Grouping factor	Effect[1]	SD
Bout ID	Intercept	<0.001
Dyad ID	Intercept	0.461
Target ID	Intercept	0.282
Target ID	Rank.diff	0.380
Target ID	CAI	0.356
Target ID	Period	0.292
Target ID	Period *CAI	0.287
Partner ID	Intercept	<0.001
Partner ID	Rank.diff	0.171
Partner ID	CAI	<0.001
Partner ID	Period	<0.001
Partner ID	Period *rank.diff	<0.001
Partner ID	Period *CAI	<0.001

Estimated standard deviations for the contribution of the random effects for the full model investigating the number of ‘favourable switches’ in the 10 s after a target individual produces an ELS, compared to the number during the remainder of dyadic grooming. CAI: composite association index.

[1] ‘Intercept’ denotes a random intercepts effect; other entries indicate a random slopes effect.

Table A16
Model 4b (requesting a favourable grooming switch; goal-directed behaviour): full model

Term	Estimate	SE	Lower CI	Upper CI	X^2	*df*	*P*	Minimum	Maximum
Intercept	−1.929	0.457	−8.994	−1.300			[1]	−8.062	−1.613
Goal met[2]	−2.959	1.119	−16.344	−1.515	15.156	1	<0.001	−18.936	−2.715
Rank.diff[3]	0.327	0.239	−0.181	1.136	2.058	1	0.151	0.197	0.463
CAI[4]	−0.045	0.246	−0.606	0.422	0.034	1	0.854	−0.238	0.186

Results of the full model investigating the probability of the target individual showing persistence, elaboration or grooming termination when they had elicited a ‘favourable’ grooming switch from the partner (goal met) or not (estimates, SEs, confidence intervals, CI, significance tests, and minimum and maximum of model estimates when dropping levels of the random effects one at a time).

[1] Not included, as has a very limited interpretation.

[2] Dummy coded with goal met=No being the reference category.

[3] z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were 3.24 and 10.33, respectively.

[4] Composite association index; z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were 2.06 and 1.32, respectively.

Table A17
Model 4b (requesting a favourable grooming switch; goal-directed behaviour): estimated standard deviations

Grouping factor	Effect[1]	SD
Bout ID	Intercept	1.435
Dyad ID	Intercept	<0.001
Target ID	Intercept	<0.001
Target ID	Rank.diff	<0.001
Target ID	CAI	<0.001
Partner ID	Intercept	<0.001
Partner ID	Rank.diff	<0.001
Partner ID	CAI	0.279

Estimated standard deviations for the contribution of the random effects for the full model investigating the probability of the target individual showing persistence, elaboration or grooming termination when they had elicited a ‘favourable’ grooming switch from the partner (goal met) or not. CAI: composite association index.

[1] ‘Intercept’ denotes a random intercepts effect; other entries indicate a random slopes effect.

Table A19
Model 5 (social anxiety): estimated standard deviations

Grouping factor	Effect[1]	SD
Bout ID	Intercept	0.805
Dyad ID	Intercept	<0.001
Target ID	Intercept	0.708
Target ID	Rank.diff	<0.001
Target ID	CAI	<0.001
Target ID	Rank.diff*CAI	<0.001
Partner ID	Intercept	0.531
Partner ID	Rank.diff	0.148
Partner ID	CAI	<0.001
Partner ID	Rank.diff*CAI	<0.001

Estimated standard deviations for the contribution of the random effects for the full model investigating the number of ELSs produced by the target individual. CAI: composite association index.

[1] ‘Intercept’ denotes a random intercepts effect; other entries indicate a random slopes effect.

Table A18
Model 5 (social anxiety): full model

Term	Estimate	SE	Lower CI	Upper CI	Minimum	Maximum
Intercept	−7.154	0.178	−7.520	−6.838	−7.213	−7.077
Rank.diff[1]	0.083	0.091	−0.082	0.256	0.034	0.142
CAI[2]	−0.039	0.086	0.215	0.136	−0.087	0.006
CAI[2]*rank.diff[1]	0.089	0.078	−0.061	0.249	0.059	0.136

Results of the full model investigating the number of ELSs produced by the target individual (estimates, SEs, confidence intervals, CI, and minimum and maximum of model estimates when dropping levels of the random effects one at a time).

[1] z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were zero and 10.94, respectively.

[2] Composite association index; z-transformed to a mean of zero and standard deviation of one; mean and standard deviation of the original variable were 2.02 and 1.31, respectively.

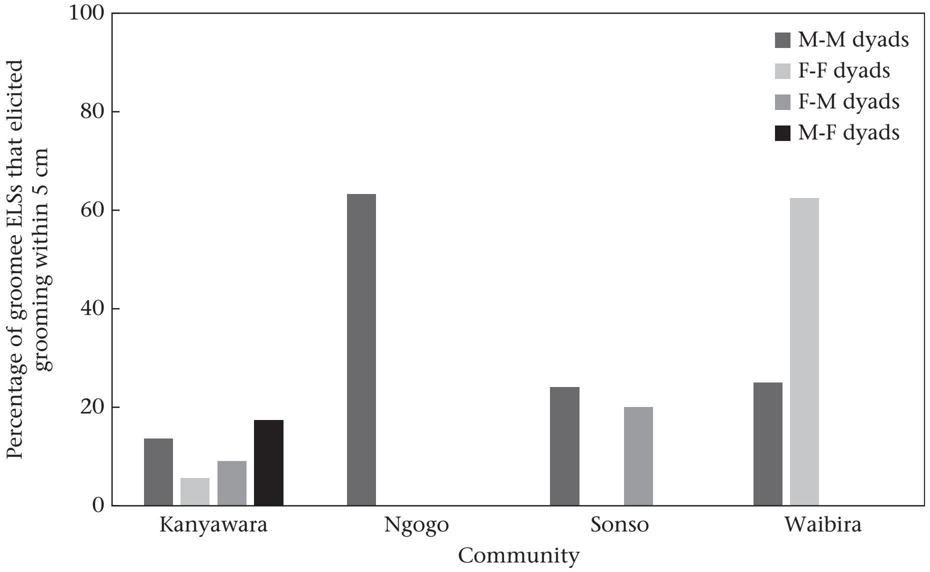


Figure A1. The percentage of groomee ELSs that elicited grooming within 5 cm of the scratched location, within 10 s after the end of the ELS. F–F dyads: female–female; M–M dyads: male–male; F–M dyads: the female was the signaller and the male the partner; M–F: dyads the male was the signaller and the female the partner. A minimum of five ELSs had to be available for a dyad class to be presented for a given community. Kanyawara (*N*=59 groomee ELSs for M–M dyads, *N*=18 F–F, *N*=22 F–M, *N*=46 M–F), Ngogo (*N*=188 M–M), Sonso (*N*=29 M–M, *N*=5 F–M), Waibira (*N*=16 M–M, *N*=8 F–F).