Behavioral Ecology and Sociobiology (2026) 80:119
https://doi.org/10.1007/s00265-026-03796-4

RESEARCH

Inferential learning explains association between self-relevance cues and overlap in chimpanzee visual bodily gesture repertoires

Anna Ilona Roberts[1,2] · **Piotr Sorokowski**[1] · **Sam George Bradley Roberts**[3]

Received: 22 January 2026 / Revised: 17 August 2026 / Accepted: 28 August 2026

Abstract

We investigated whether the association of self-relevance cues with overlap in visual bodily gestures is better explained by inferential learning or experience-based reinforcement of gesture–outcome contingencies. Inferential accounts predict that self-relevance cues facilitate convergence by supporting goal inference in complex socio-ecological situations, in secondary contexts and responses. In contrast, experience-based accounts predict that self-relevance cues facilitate repeated exposure, promoting convergence through reinforcement of stable gesture–outcome associations in primary contexts. Using generalized linear mixed models and social network analyses of 12 wild East African chimpanzees, we examined relationships among self-relevance cues, social relationships, socio-ecological conditions, contextual specificity, and gestural repertoire overlap. Overall, the results are more consistent with inferential accounts than reinforcement accounts. Although reciprocated social bonds provided opportunities for reinforcement learning, they did not predict repertoire overlap, indicating exposure is insufficient for convergence. Three findings supported inferential account. First, self-relevance cues accompanied heterogeneous gestures in socially complex and uncertain situations, including interactions with non-kin, larger parties, and weaker bonds. Second, self-relevance cues such as vocalizations, manual gestures, and response waiting were associated with secondary contexts and responses rather than with primary contexts expected under reinforcement learning. Third, repertoire convergence was predicted by gestures in secondary contexts, whereas gestures in primary contexts did not predict convergence. Consistent with inferential account, these findings suggest that self-relevance cues may facilitate inference of communicative goals under uncertainty, potentially supporting flexible alignment of gestural repertoires in dynamic fission–fusion societies, explaining how gestural convergence and divergence coexist, making gesture culture likely.

Significance statement

Understanding how communicative repertoires converge among dyad partners is central to explaining the evolution of social complexity. We tested whether overlap in wild chimpanzee visual bodily gestures is better explained by experience-based reinforcement of gesture–outcome associations, or by inferential processes to interpret communicative goals. Analyses of 12 wild chimpanzees showed that self-relevance cues accompanied heterogeneous gestures in complex social conditions, predicted secondary rather than primary communicative contexts, and were linked to greater repertoire overlap only when gestures occurred in secondary contexts. Opportunities for reinforcement learning through reciprocated social bonds did not predict convergence. These findings are more consistent with the interpretation that self-relevance cues facilitate inference of communicative goals than simply increasing exposure to signals and outcomes. Communicative convergence in fission–fusion societies may emerge through flexible goal attribution under uncertainty, providing a mechanism through which socially complex species can maintain divergence and convergence in gestural communication.

Keywords Chimpanzee · Gestural communication · Inferential learning · Trial-and-error learning · Self-relevance cues · Repertoire overlap

Communicated by K. Strier

Extended author information available on the last page of the article

Published online: 05 September 2026

Introduction

Navigating complex social relationships requires individuals to communicate successfully with partners who differ in experience, social history, and expectations (Dunbar 2025). To achieve this, primates rely heavily on self-relevance cues—audience-directed behaviours that explicitly indicate to a recipient that another individual is interacting with them (Roberts and Roberts 2018; Sperber and Wilson 2026). In primate communication, some signals carry propositional content (Tomasello et al. 2019) but are not always targeted effectively and may lack a clear recipient (Amici et al. 2025). Consequently, signallers use self-relevance cues, including observable behaviours such as direct visual attention, indicative manual gestures (e.g., directed hand-reaches) (Lyn et al. 2014) and vocalisations to establish communicative intent and influence partner's attention (Roberts et al. 2022a; Gomez 2026). While these directed cues are known to influence how signals are processed, the precise cognitive mechanisms underlying their effects remain poorly understood.

Gestural communication varies in complexity which may influence the ability of the recipient to respond effectively to signals. This complexity is particularly evident in visual bodily signals—defined here as voluntary movements of the body, legs, posture and locomotory gaits (Hewes 1973; Pika and Tomasello 2002; Liebal et al. 2004; van Boekholt et al. 2024). Visual bodily gestures are intentional, meaning signaller has a goal and some flexibility in attaining it (Pika and Tomasello 2002). However, not all visual bodily signals are equally flexible. Their degree of behavioural flexibility varies with the signaller's emotional arousal (Aureli et al. 1999; Behringer et al. 2012; Barrault et al. 2022). Under high arousal (e.g., intense agonistic threats), behavioural outputs are often highly constrained and predictable, driving rapid, stereotyped responses from recipients (Zajonc and Sales 1966). Conversely, low-intensity visual bodily signals are frequently detached from high physiological arousal. This independence grants them a high degree of flexibility; the same bodily signal can occur across diverse contexts with varied underlying motivations. Consequently, these low-intensity signals exhibit greater contextual ambiguity (Roberts and Roberts 2016a).

This raises a question about mechanisms of resolution of ambiguity in visual bodily gestures. One such potential mechanism is self-relevance cues which frequently accompany or precede visual bodily gestures. Use of self-relevance cues in conjunction with visual bodily gestures is associated with greater specificity of responses by the recipient (Mine et al. 2024; Nellissen et al. 2026). Further, self-relevance cues are associated with greater overlap in visual bodily gesture repertoires between individuals (Roberts and Roberts 2022). These associations have led researchers to propose that resolution of ambiguity plays important role in the acquisition of repertoire of visual bodily signals by the recipient (Plooij 1978; Tomasello 1999; Laland and Janik 2006; van Schaik et al. 2012; Whiten 2021; Roberts and Roberts 2022). However, the cognitive mechanisms underlying role of self-relevance cues in overlap of visual bodily gestures remain unclear. Using data from wild chimpanzees (*Pan troglodytes schweinfurthii*), we test whether self-relevance cues are associated with variation in visual bodily gesture overlap consistent with inferential goal attribution, or with experience-based reinforcement of gesture–outcome contingencies.

To resolve ambiguity in visual bodily gestures, recipients can deploy distinct cognitive strategies. One solution is the experience-based reinforcement of signal–outcome contingencies, as proposed by Ontogenetic Ritualization and the Social Negotiation Hypothesis (Plooij 1978; Tomasello et al. 1994; Hobaiter and Byrne 2011; Pika and Fröhlich 2019; Fröhlich et al. 2025b). These accounts propose that gestures emerge through direct experience of gesture–outcome contingencies, with proximity and visual attention increasing opportunities for repeated learning (Coussi-Korbel and Fragaszy 1995). Over time, repeated learning via trial and error strengthens associations between gestures and outcomes, producing convergence in repertoires. Particularly, strongly bonded dyads should develop a similar repertoire of gestures because animals form the same associations between signals and outcomes through interacting across similar contexts.

However, reliance on trial and error may reduce the efficiency of social interactions in dynamic social environments where interaction partners and contexts frequently change (Roberts et al. 2022b). In fission–fusion societies, large number of group members and frequent changes in subgroup composition reduce opportunities for repeated interactions (Dunbar 1998). As social and ecological complexity increases, the need for more sophisticated, diverse, and flexible communication systems to navigate complex social dynamics may also increase (Freeberg et al. 2012; Roberts and Roberts 2020; Pedruzzi et al. 2026). Communicative complexity, defined as the diversity of communicative signals and the information they transfer may facilitate this process by enabling flexible coordination of social relationships across diverse social partners (Freeberg et al. 2012).

In line with this challenge of maintaining relationships across multiple and diverse social partners, the Communicative Roots of Complex Sociality and Cognition Hypothesis proposes that representing the goals underlying signal production facilitates the formation and maintenance of social bonds more effectively than less cognitively complex

forms of communication (Roberts and Roberts 2020). This advantage should be particularly important among weakly bonded or infrequently encountered partners who must infer the goals underlying visual bodily gestures in complex social settings (Dunbar 1998). These inferences demand that recipients integrate contextual information, including the broader social context, the identity of the signaller, and the relationship between signaller and the recipient (Roberts and Roberts 2016a; Schleihauf et al. 2025). Inferential processes that track communicative goals of weakly bonded social partners in these social settings may generate both convergence and divergence in gestural repertoires (Watson et al. 2014; Roberts 2025). Convergence emerges when individuals infer shared goals and social identities underlying gestures, whereas divergence arises across dyads and groups who differ in goals and identities (Tomasello et al. 2005; Roberts and Roberts 2026).

These distinct learning strategies make different predictions regarding the role of self-relevance cues in modulating responses to and overlap in visual bodily gestures. Visual bodily gestures can be characterized by their social context and functional outcome. Context refers to the immediate behavioural situation (e.g. grooming, travel, feeding, or play), whereas function refers to the social outcome of the interaction, such as affiliative, offensive, or defensive (Roberts et al. 2012a; Fröhlich et al. 2025a). Visual bodily gestures may occur in association with the most common, primary contexts or responses, or in association with the less common, secondary contexts or responses (Roberts and Roberts 2022).

The experience-based accounts predict the strongest associations between self-relevance cues and primary contexts or responses. If self-relevance cues consistently accompany visual bodily gestures, they become incorporated into the learned communicative cue complex through repeated co-occurrence, strengthening retrieval of the primary gesture–outcome association. Under this account, neither signallers nor recipients acquire gestures through representing the goals underlying social interactions (Bouton 2021). Repeated learning via trial and error would instead promote convergence among individuals frequently interacting in similar contexts through direct reinforcement of primary gesture–outcome associations.

In contrast, inferential accounts predict that self-relevance cues should be most strongly associated with secondary contexts, and responses, potentially operating within broader multimodal communicative strategies involving vocalisations, gestures, and facial expressions (Vick et al. 2007; Leavens et al. 2010; Taglialatela et al. 2015; Wilke et al. 2017; Mine et al. 2024; van Boekholt et al. 2024; Nellissen et al. 2026). In this framework, self-relevance cues enable flexible allocation of attention between dominant (primary) and weaker (secondary) representations (Webb and Graziano 2015). Under this account, self-relevance cues support the shifts of attention away from habitual response tendencies, thereby facilitating inference about the signaler's likely goals under conditions of uncertainty (Webb et al. 2016b). Successful communication may therefore depend increasingly on self-relevance cues, allowing recipients to integrate contextual information and infer secondary representations underlying visual bodily gestures (Roberts and Roberts 2018; Cartmill 2023). In this inferential framework, recipients acquire gestures through representing the goals underlying social interactions. In this case, convergence in visual bodily gestures would depend on goal inference beyond direct experience of primary contexts or responses in situations where individuals share mutually desired social goals (Baron-Cohen 1994; Roberts and Roberts 2015).

Socio-ecological conditions provide a further means of distinguishing these competing hypotheses. Successful communication depends not only on the characteristics of the interacting dyad, such as kinship, age and sex similarity, reproductive state, proximity and relationship quality, but also on broader socio-ecological conditions, including audience composition, party size, visibility, illumination, noise and human presence (Shultz and Dunbar 2022). These factors influence both the amount of competing social information available to recipients and the uncertainty surrounding communicative intent (Roberts 2025; Pedruzzi et al. 2026). Experience-based accounts posit that unpredictability biases animals to interpret ambiguous signals negatively (Payne et al. 2007; Bethell et al. 2012; Davis et al. 2016; Roelofs et al. 2016). Consequently, experience-based accounts would predict that in complex socio-ecological conditions, the effects of self-relevance cues should reinforce these associations leading to avoidance and reduced overlap (Bouton 2021).

In contrast, inferential accounts predict that increasing socio-ecological complexity increases uncertainty about goals of social partners, potentially leading signallers to use a more diverse or less overlapping set of visual bodily gestures across interactions and making understanding of goals underlying visual bodily gestures less reliable (Roberts 2025). Under these conditions, self-relevance cues should become increasingly important for directing attention towards behaviourally relevant information and integrating contextual information required to infer communicative goals (Roberts et al. 2022a; Sperber and Wilson 2026). Accordingly, if self-relevance cues facilitate inferential goal attribution, their association with repertoire overlap should be strongest under conditions of high socio-ecological complexity, whereas experience-based accounts predict overlap to arise primarily through repeated reinforcement of gesture–outcome contingencies in less complex conditions.

Previous research analyzed predictors of dyadic overlap in chimpanzee visual gesture repertoires, measured using kappa values, as well as predictors of the repertoire sizes of homogeneous gestures (gesture types shared between interacting individuals) and heterogeneous gestures (gesture types unique to one interaction partner) (Roberts and Roberts 2017). These analyses showed that self-relevance cues predicted overlap in chimpanzee visual bodily gesture repertoires, whereas proximity and visual attention considered in the same model did not predict greater overlap (AIR, unpubl. data). However, these findings do not explain why self-relevance cues predict repertoire overlap. Self-relevance cues may increase overlap because they reinforce primary gesture–outcome associations, or they may increase overlap by facilitating inference of communicative goals. To address this question, this study examined gestural communication in wild chimpanzees and predicted that, under inferential accounts, self-relevance cues would be used in complex socio-ecological environments and secondary contexts. In contrast, under experience-based accounts, self-relevance cues would be used in primary contexts, when socio-ecological complexity is low. Accordingly, we tested whether the association between contexts and repertoire overlap was strongest in primary contexts under predictions derived from experience-based reinforcement, or in secondary contexts, as predicted by inferential accounts.

Methods

Data were collected from 12 adult East African chimpanzees (*Pan troglodytes schweinfurthii*) of the Sonso community at the Budongo Conservation Field Station, Uganda. Data were collected during March–June 2008. Details of focal subjects and observation effort are provided in Online Resource 1, Table S1. The sample was restricted to adults only to reduce influence of potential ontogenetic variation in gesture production and usage (Tomasello et al. 1994; Liebal et al. 2004). Online Resource 2.1 specifies details of study site and population. Each chimpanzee was observed during standardized 20-minute focal follows, with sampling distributed across different times of day throughout the study period. For subsequent analyses, only the first 18 min of each focal follow were retained and aligned with behavioural observations using the socioecological scan-samples. The animals were chosen systematically for focal follows which were taken at least 20 min apart. Gestural interactions and the associated behavioural context were continuously recorded using handheld digital video camera (SONY DCR-HC18E/HC32E). Simultaneously, verbal commentary accompanied filming by the camera to document the identity of the signaller and recipient, the behavioural context before and after gestural production, and the bodily orientation of interactants relative to one another during communication events. Because precise gaze direction is difficult to determine reliably under natural forest conditions, bodily orientation was used as a proxy for visual attention (see Online Resource 2.2 for details of bodily orientation recording). This is a widely used indicator of attentional orientation in field studies of primate communication (Bard 1992; Tomasello et al. 1994), although it does not provide a direct measure of gaze. Previous analyses of this chimpanzee community showed that in 91.1% of adult-to-adult gestural interactions, signallers were bodily oriented towards recipients during gesture production (Roberts and Roberts 2019), supporting the reliability of this measure for identifying recipient-directed communication while also acknowledging that orientation may overestimate actual gaze alignment in some cases. Vocalisations occurring during or immediately surrounding gestural events were recorded continuously using the video camera, allowing assessment of whether gestures were produced with or without accompanying vocal signals.

Focal follows were paired with scan sampling of the surrounding social and ecological environment. Each focal follow consisted of nine scans recorded at 2-minute intervals, recording: (1) the identity of all individuals located within 10 m of the focal subject, as well as individuals located further than 10 m away within the same party; (2) the identity of the focal individual's nearest adult neighbour; (3) the distance between the focal subject and nearest neighbour, measured in metres; (4) the bodily orientation of the focal individual relative to the nearest neighbour, and vice versa; and (5) visual access between the focal subject and the nearest neighbour. Further, the activity and height above ground level in meters of the focal subject and the nearest neighbour were recorded. A complete description of the variables recorded is presented in Online Resource 2.3.

Ecological variables were sampled concurrently at 6-minute intervals across the same 18-minute period. These scans recorded ambient noise levels, illumination, wind conditions, visibility, visitor number, and visitor distance. Temperature was recorded at the end of each focal sample. A complete description of ecological data collection procedures is provided in Online Resource 2.4. Two experienced field assistants recorded the socio-ecological data for this study checking its reliability on weakly basis. Annual inter-observer reliability test of group composition is conducted across field assistants to ensure the consistency of scoring, with results consistently above 0.85 Spearman's rank correlation coefficient, r_s. Spearman's rank correlation was used because the measures of group composition and association patterns were based on ordinal and proportional observational data that did not necessarily meet assumptions of normality required for parametric tests. Spearman's r_s is widely

used in observational studies to assess the consistency of ranked or frequency-based measures between observers (Bateson and Martin 2021). High coefficients indicated strong agreement between observers in recording the presence and association patterns of individuals within parties. It was not possible to record data blind because our study involved focal animals in the field, however field assistants recording the socio-ecological data for this study were unfamiliar with the aims of the research.

Video analyses of communication

First, a catalogue of video recordings of nonverbal behaviors was compiled from the video footage. In this study, "nonverbal behaviors" refers to visible bodily actions, including movements of the limbs, head, torso, body postures, and locomotory gaits that function communicatively. A complete description of the classification of behavior as gestures is provided in Online Resource 2.5. An ethogram of Budongo gestural communication is provided in Online Resource 3 for clarity and reproducibility (Roberts et al. 2012b, 2014).

To classify gestures according to social context, we examined the eliciting social or ecological conditions occurring immediately before, and after each gestural event, scoring each event against following criteria (Roberts and Roberts 2016b): (1) courtship (behaviour where adult male maintains monopoly of their sexual partner), (2) food (sharing, eating or observing others eat plant food or meat), (3) groom (using thumb or index finger to push through hair on another's body to pick at exposed skin, groom initiations), (4) inter-community (interactions in context of hearing other communities or patrolling their territory), (5) reunion (communicating or interacting in context of hearing another party or meeting after separation), (6) third-party aggression (observing aggressive behaviours between two other individuals), (7) play (bodily contact or interacting in a non-agonistic manner such as wrestling, chasing or tickling), (8) sex (mounting, stimulating genitals, copulating), (9) travel (walking, running with or following another in certain direction).

Gestures were characterized according to function: (1) affiliative if they lead to increased cohesion between signaler and the recipient (e.g. initiation of grooming), (2) defensive (e.g. appeasement in response to receiving or observing aggression, including reconciliation and reassurance behaviors) and (3) offensive (e.g. aggressive behavior leading to physical aggression, retaliation, etc.). The response occurred without fixed time window when the change in recipient's behavior occurred immediately after the gesture or alternatively when immediate response was absent.

Second coder coded a random sample of 45 gesture sequences for similarity in the classification of gesture function and sensory modality, showing good reliability for function (Cohen's $K = 0.70$) and excellent reliability for modality ($K = 0.946$). This reliability sample represented 1.86% of the full dataset comprising 2417 gesture sequences recorded across 12 focal subjects during three study periods. Although this proportion is smaller than the 10% reliability samples often discussed in some communication studies, it is consistent with studies of gestures due to the substantial time required for frame-by-frame behavioural coding (e.g. Hobaiter and Byrne 2011). The reliability sample was selected randomly to ensure representation of the broader dataset and to evaluate consistency in the application of coding definitions rather than to recode the entire dataset. The obtained Cohen's Kappa values indicate substantial agreement according to established benchmarks for reliability (Landis and Koch 1977; Bakeman and Gottman 1997). An additional random sample of 50 gesture sequences was independently coded for response waiting and persistence, yielding good reliability (Cohen's $K = 0.74$) (Bakeman and Gottman 1997).

Cohen's kappa coefficient

In this study we examined associations between dyadic repertoire overlap (κ) and measures of social, ecological, and communicative complexity. Following previous research (Tomasello et al. 1997; Genty et al. 2009), we quantified gestural repertoire overlap using Cohen's kappa, based on a binary presence–absence matrix in which each gesture type was scored as present or absent for each individual within a dyad. Kappa values were used as a measure of agreement between dyad members, where $\kappa = 1$ indicates complete overlap and $\kappa = -1$ indicates complete divergence. Analyses were restricted to visual bodily gestures from adult–adult interactions collected across three study seasons.

Analyses

1) Non parametric statistics.

Gesture specificity to context, function, and response was calculated as the mean proportion of cases in which a gesture type co-occurred with its primary context, affiliative function, or recipient response (Table 1; Online Resource 1, Table S2). Gestures often co-occur with one context, function and response type more frequently than other contexts, functions and responses. The most commonly seen context, function and response type for each gesture type was labelled as primary context, function or response. All events were then recategorized as dichotomous variable: either as primary (matching most frequent context, function

Table 1 Specificity to primary and secondary contexts across gesture types

Gesture type	Specificity to primary context	Primary context	Secondary contexts (in brackets) and their specificity
bob	38.5	reunion	23.1 (food), 7.7 (inter community), 7.7 (observe third party aggression), 23.1 (travel)
bow	47.8	travel	8.7 (observe third party aggression), 34.8 (reunion), 8.7 (sex)
crouch	36.8	reunion	10.5 (food), 10.5 (groom), 5.3 (inter community), 5.3 (sex), 31.6 (travel)
crouch run	42.4	reunion	3.4 (food), 5.1 (groom), 5.1 (inter community), 6.8 (observe third party aggression), 15.3 (sex), 22.0 (travel)
crouch walk	37.9	reunion	1.7 (courtship), 10.3 (food), 3.4 (groom), 5.2 (inter community), 3.4 (observe third party aggression), 5.2 (sex), 32.8 (travel)
dangle	83.3	travel	16.7 (courtship)
drag self	100.0	travel	-
hang	100.0	reunion	-
hold object	100.0	sex	-
jump	42.9	travel	14.3 (food), 14.3 (reunion), 28.6 (sex)
lower head	85.7	groom	14.3 (food)
lunge	50.0	food	16.7 (observe third party aggression), 16.7 (sex), 16.7 (travel)
nod	100.0	reunion	-
present genitals	100.0	sex	-
present leg	100.0	groom	-
present mount	100.0	sex	-
present rump	26.1	reunion, travel	4.3 (courtship),8.7 (food), 4.3 (groom), 17.4 (inter community), 13.0 (sex)
present torso	84.5	groom	8.6 (inter community), 5.2 (reunion), 1.7 (travel)
rock	66.7	reunion	33.3 (travel)
roll over	100.0	groom	-
run stiff	40.0	reunion, travel	10.0 (courtship), 2.5 (food), 2.5 (inter community), 5.0 (play)
sniff	100.0	travel	-
stationary stiff	53.8	courtship	7.7 (inter community), 7.7 (reunion), 7.7 (sex), 23.1 (travel)
swagger bipedal	70.0	travel	10.0 (food), 20.0 (reunion),
swagger quadrupedal	80.0	travel	20.0 (reunion)
sway	75.0	travel	25.0 (food)
swing	42.9	sex	14.3 (groom), 14.3 (reunion), 28.6 (travel)
tip head	100.0	sex	-
turn back	42.9	reunion	14.3 (food), 14.3 (sex), 28.6 (travel)
turn head	100.0	reunion	-
walk stiff	29.4	travel	17.6 (courtship), 29.4 (reunion), 23.5 (sex)

Calculated for focal adult to adult gesture events only. Primary context type was calculated as a percentage of gesture events pulled across all subjects occurring within the most common context type

or response) or secondary that did not match the primary context, function or response assigned to a gesture type (Roberts et al. 2012a; Fröhlich et al. 2025a). These variables were derived from observed behavioural distributions and therefore represent a descriptive measure rather than an independent predictor. Whilst this does not affect the validity of the statistical models, it does warrant caution in causal interpretation of effects involving this measure. Differences in specificity were assessed using two-tailed Wilcoxon signed-ranks tests ($\alpha = 0.05$). Medians and interquartile ranges (IQ) are reported.

2) Generalized linear mixed models.

First, we report results from Generalised Linear Mixed Models (GLMMs) examining the influence of social and ecological variables on dyadic gestural overlap (κ). Variable definitions, coding schemes, and statistical results are provided in the Online Resources. The first GLMMs (Online Resource 4; data in Online Resource 11) included 21 social, ecological, and communicative predictors of kappa values

between dyad partners, including nearest-neighbour variables, environmental conditions, group composition, and gestural production rates. Nearest-neighbour variables were included given evidence that chimpanzees preferentially gesture towards the nearest individual (Roberts and Roberts 2015). Second, we examined associations between self-relevance cues and the likelihood of primary context and response types (Online Resource 5; data and results in Online Resource 12 and Online Resource 6). Analyses were conducted separately for homogenous and heterogeneous gestures and included both single and non-sequential events. Predictors included self-relevance cues (e.g. mutual visual contact, vocalisations, manual gestures) and social bonds based on grooming reciprocity, with controls for gesture frequency and social relationship measures. Third, we tested whether social and ecological variables predicted the occurrence of homogenous versus heterogeneous gestures (Online Resource 7; data in Online Resource 13). Initial models included 12 predictors (Table S2), with a second step adding four additional self-relevance variables (Table S3), including mutual visual contact, vocalisations, and manual gesture presence. Finally, we conducted separate GLMMs for homogenous and heterogeneous gestures to examine predictors of primary context (Online Resource 8; data in Online Resource 14). For heterogeneous gestures, predictors included age composition of partners, grooming reciprocity, and accompanying vocalisations and manual gestures. For homogenous gestures, predictors included age composition, grooming reciprocity, and mutual visual contact prior to gestures (Tables S2–S3).

The gestural data has a hierarchical structure in which gestures are clustered by focal individuals and directed at specific individuals. The GLMMs were therefore structured with Level 1 as the identity of the focal individual/signaller and Level 2 as the identity of the recipient of the gesture or identity of the nearest neighbour. The models were fitted using a binomial error structure with logit link. To account for the hierarchical structure of the data, the GLMMs included random effects for the focal individual identity (to account for multiple gestures from the 12 focal individuals) and focal individual identity by recipient individual (to account for multiple gestures from a specific focal individual directed towards a specific recipient). To limit the complexity of the model and ensure model convergence, random intercepts but not random slopes were used for the random effects. Analyses were conducted in SPSS 25.0. SPSS does not produce Variance Inflation Statistics (VIF) for GLMMs. Therefore, multicollinearity was assessed using linear regression to obtain VIFs, with the same outcome and predictor variables as those used in the GLMMs. For the models presented in Online Resource 4, Table S2, and Online Resource 8, Tables S2 and S3 all VIF values were below 5, indicating no issues with multicollinearity. For the models in Online Resource 6, Table S2 (max VIF value 7.17), Online Resource 7, Table S2 (max VIF value 6.53) and Online Resource 7, Table S3 (max VIF value 6.6), the VIF values suggest the possibility of moderate multicollinearity. However, across all the models, no VIF values exceeded 10, suggesting that multicollinearity is unlikely to be a serious issue with the GLMMs (Kutner et al. 2005).

3) Social network analysis.

First, we used social network analysis to examine associations between repertoire overlap and rates of social interaction (interaction contexts and frequencies). Second, we tested whether rates of gestural communication in primary and secondary contexts predicted overlap in visual bodily gestures, and whether secondary-context rates were associated with primary-context rates. Details of these analyses are provided in Online Resource 9, Tables S4 and S5.

Demographic and relational factors (age, kinship, sex, and reproductive state) were included in all models. Dyads were classified as same age class (≤5 years difference) or different age class (>5 years) (Mitani et al. 2002; Online Resource 9, Table S1). Female reproductive status was defined based on sexual swelling, with females classified as oestrous when maximal tumescence and mating were observed; all focal males were reproductively active. Sex similarity was coded as same- or mixed-sex dyads. Further details of attribute coding are provided in Online Resource 9, Table S1.

We also used social network analysis to examine relationships between gestural homogeneity and function (Online Resource 9, Table S2). Six behavioural networks (joint feeding, travel, resting, grooming given, grooming received, and mutual grooming per hour in party) were related to repertoire homogeneity (Online Resource 9, Table S3; Online Resource 16). All matrices were weighted and directed, with cell values representing interaction rates for each dyad and direction. MRQAP (Double Dekker semi-partialling) was used to test relationships between behavioural networks. Degree centrality (in- and out-degree) was computed from normalized matrices (Croft et al. 2010), and all transformations were conducted in UCINET 6 (Borgatti et al. 2014).

Results

Specificity of bodily, visual signals to context, function and response

Chimpanzees produced homogeneous visual bodily gestures across a median of 1.61 contexts (IQ = 1.26–1.83) and heterogeneous gestures across 1 context (IQ = 1.00–1.51),

with maxima of 5 and 4 contexts, respectively. Homogeneous gestures occurred most frequently in reunion (median proportion = 0.30, IQ = 0.11–0.51), travel (0.27, IQ = 0.02–0.39), and grooming contexts (0.14, IQ = 0.03–0.32). Heterogeneous gestures were also most common in reunion (0.30, IQ = 0.06–0.79), followed by travel (0.15, IQ = 0.00–0.37) and sex contexts (0.07, IQ = 0.00–0.54).

Both homogeneous (median primary-context specificity = 0.80, IQ = 0.78–0.83) and heterogeneous gestures (1.00, IQ = 0.86–1.00) showed high primary-context specificity, although specificity was greater for heterogeneous gestures (Z = -2.090, N = 10, P = 0.037; Table 1).

When classified by function, both homogeneous (median affiliative specificity = 0.43, IQ = 0.31–0.63) and heterogeneous gestures (0.12, IQ = 0.00–0.27) were associated with affiliative outcomes, although this association was stronger for homogeneous gestures (Z = -2.666, N = 10 [1 tie], P = 0.004).

Finally, both homogeneous (median = 0.52, IQ = 0.30–0.58; Z = -2.903, N = 12, P = 0.001) and heterogeneous gestures (median = 1.00, IQ = 0.81–1.00; Z = -2.507, N = 9, P = 0.012) elicited recipient responses, although responses were more likely following heterogeneous gestures (Z = -2.547, N = 9, P = 0.008).

Influence of social and ecological variables on overlap in repertoire of visual, bodily gestures

A greater overlap (kappa value) in visual bodily gesture repertoires between focal subjects and their nearest neighbours occurred when partners were of the same age category, non-kin, male, positioned lower above the ground, and mutually oriented towards one another (See Online Resource 4; data in Online Resource 11). Overlap was also greater under higher illumination and visibility, when fewer same-age partners were present within 10 m, and when vocalisations and tactile manual gestures occurred more frequently.

Association of self-relevance cues with primary context and primary response to visual bodily gesture

We examined whether primary contexts and primary responses to visual bodily gestures were associated with reciprocated grooming between signaller and recipient. For heterogeneous gestures, unreciprocated grooming was associated with a higher likelihood of both a primary response (β = 2.575, p = 0.006) and occurrence in a primary context (β = 1.418, p = 0.035). For homogeneous gestures, reciprocated grooming was associated with a higher likelihood of a primary response (β = -1.052, p = 0.006), but not a primary context (β = 0.682, p = 0.166).

We then tested whether self-relevance cues predicted primary contexts (Fig. 1) and primary responses (Online Resource 10, Fig. S1), while controlling for grooming relationships and the frequency of homogeneous and heterogeneous gestures within sequences. Vocalisations during gesturing, tactile manual gestures, and response waiting were associated with a higher likelihood of secondary responses. Similarly, manual visual gestures, vocalisations prior to gesturing, and response waiting were associated with a higher likelihood of secondary contexts. In contrast, mutual visual contact and tactile manual gestures were associated with a higher likelihood of gestures occurring in primary contexts.

Influence of social and ecological variables on the use of homogenous and heterogeneous visual, bodily gestures

We used GLMMs to examine predictors of heterogeneous relative to homogeneous visual bodily gestures. Initially, chimpanzees were more likely to use heterogeneous gestures in larger parties, when interacting with non-kin, when grooming reciprocity was present, when joint resting and travel durations were shorter, and when fewer same-age partners as the recipient were present nearby (see Online Resource 7; data in Online Resource 13). After including self-relevance cues, heterogeneous gestures remained associated with larger parties, non-kin and reproductively active partners, shorter joint resting and travel, absence of mutual visual contact before gesturing, and the presence of vocalisations and manual visual gestures.

We then examined predictors of primary versus secondary contexts. For heterogeneous gestures, a larger number of same-age partners as the recipient predicted primary contexts (β = 12.266, p = 0.001), whereas manual visual gestures predicted secondary contexts (β = 2.592, p = 0.024). Vocalisations showed a trend towards predicting secondary contexts (β = 1.032, p = 0.071). For homogeneous gestures, neither audience composition, grooming reciprocity, nor mutual visual contact prior to gesturing significantly predicted context type.

Association between the rate of production of gestures in primary and secondary contexts and repertoire overlap in visual bodily signals

We used MRQAP to examine associations between rates of gesture production in primary and secondary contexts (events/hour within 10 m) and dyadic homogeneity of visual bodily gestures (kappa value). Demographic variables (maternal kinship, sex similarity, age similarity, and oestrous similarity) were included as controls. Dyads

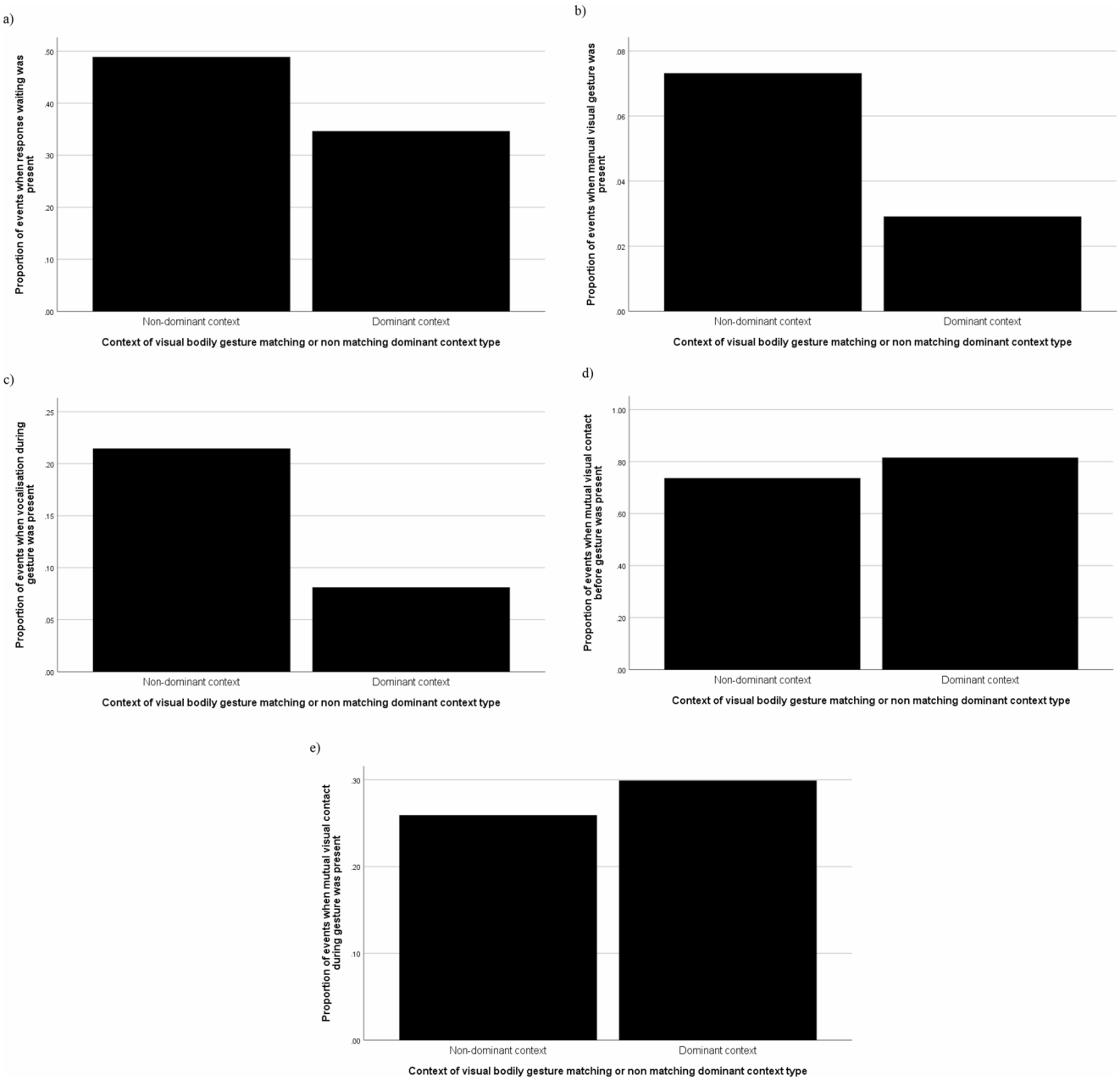

Fig. 1 Predicted probability of context matching or not matching primary context for visual bodily gesture type as a function of self-relevance cues (**a**–**e**). Panels show response waiting, manual visual gestures, vocalisations, and mutual visual attention before and during gesturing

producing gestures in secondary contexts at higher rates showed greater gestural homogeneity ($r^2 = 0.426$, $\beta = 0.099$, $p = 0.049$), whereas production of gestures in primary contexts was unrelated to homogeneity ($\beta = 0.096$, $p = 0.143$). Dyads producing more gestures in secondary contexts also produced more gestures in primary contexts ($r^2 = 0.202$, $\beta = 0.378$, $p = 0.016$).

Association of repertoire overlap with rate of social interaction and centrality in the network

We examined associations between visual bodily gesture complexity (gesture rate, social bonding duration, and frequency of gestures by function) and dyadic homogeneity (kappa value), controlling for maternal kinship, sex similarity, age similarity, and oestrous similarity. For brevity, only significant results are reported.

MRQAP analyses showed that dyads producing visual bodily gestures at higher rates exhibited greater gestural

homogeneity ($r^2 = 0.429$, $\beta = 0.168$, $p = 0.035$). Homogeneity was positively associated with same-sex dyads ($r^2 = 0.515$, $\beta = 0.434$, $p = 0.002$), synchronized low-intensity pant-hoots ($\beta = 0.193$, $p = 0.001$), reassurance ($\beta = 1.213$, $p = 0.013$), and other threat behaviours ($\beta = 0.107$, $p = 0.049$), and negatively associated with mutual grooming ($\beta = -0.764$, $p = 0.030$).

We then examined associations between six social networks (joint feeding, travel, resting, grooming given, grooming received, and mutual grooming) and repertoire homogeneity. Of these, only joint travel predicted greater gestural homogeneity ($r^2 = 0.441$, $\beta = 0.254$, $p = 0.003$).

Node-level analyses showed that individuals receiving more grooming ($r^2 = 0.677$, $\beta = 0.710$, $p = 0.038$) and engaging in more joint travel ($r^2 = 0.742$, $\beta = 0.964$, $p = 0.013$), but giving less grooming ($\beta = -0.795$, $p = 0.021$), exhibited greater overlap in visual bodily gestures.

Discussion

The aim of this research was to examine whether the association of self-relevance cues with overlap between individuals in their repertoire of visual bodily gestures is better explained by inferential goal attribution, or by experience-based reinforcement of gesture–outcome contingencies (Roberts et al. 2022b). Taken together, results of analyses of self-relevance cues, social relationships, and socio-ecological conditions in relation to the use, contextual specificity, and repertoire overlap of gestural communication are more consistent with theoretical accounts invoking inferential goal attribution than with models based on direct experience of gesture–outcome contingencies (Fröhlich et al. 2025b).

A central prediction of experience-based accounts is that self-relevance cues should facilitate repeated experience of stable gesture–outcome contingencies. Under this account, recipients are better able to retrieve the primary, most common outcome accompanying use of visual bodily gestures facilitating repertoire overlap (Carrasco et al. 2004; Noudoost and Moore 2011). If reinforcement is the main mechanism underlying repertoire overlap, self-relevance cues should be most strongly associated with primary contexts and primary responses. The present findings provide limited support for this prediction.

Some self-relevance cues, such as mutual visual contact during gesturing, predicted primary contexts, suggesting that direct visual contact facilitates responding in most common context types for a given visual bodily gesture type. However, these effects were not the strongest predictors of signalling overlap between individuals. Whilst mutual visual contact alone was associated with greater overlap in visual bodily gesture repertoires in a previous study, this relationship was no longer evident when mutual visual contact was considered alongside other self-relevance cues. Indeed, after accounting for accompanying manual gestures and vocalisations, greater mutual visual contact was associated with a smaller repertoire of overlapping visual bodily gestures.

Self-relevance cues may facilitate inferential processing by being used in situations where recipients prioritize weaker or secondary representations over more immediate primary responses (Roberts and Roberts 2022). In line with this prediction, the use of self-relevance cues was primarily associated with secondary responses and secondary contexts. Vocalisations during gestures, tactile manual gestures, and response waiting predicted secondary responses, while vocalisations prior to gesturing, manual visual gestures, and response waiting predicted secondary contexts. These findings are difficult to reconcile with a simple reinforcement account because secondary contexts and responses, by definition, occur less frequently and therefore the outcome cannot be derived solely from the features of the signals associated with most common outcome.

It could be argued that these self-relevance cues co-occur with secondary contexts because ongoing social cues resolve ambiguity by retrieving previous secondary gesture-outcome associations (Bouton 1994). In line with this possibility, we found that across heterogeneous visual bodily gestures, secondary response type and secondary context were predicted by the presence of reciprocated social bonds, consistent with interpretation that previous reinforcement history influenced these interactions.

However, reciprocated bonds, which reflect high proximity and visual attention (Roberts and Roberts 2016b; Lewis et al. 2021) should facilitate repertoire overlap, as individuals in closer and more attentive relationships should be more likely to experience similar gestures and outcomes, leading to greater repertoire overlap (Hobaiter and Byrne 2011; Pika and Fröhlich 2019). As reciprocated bonds predicted secondary contexts across heterogeneous gestures (un-overlapping within dyads), our findings are inconsistent with this suggestion. Although contextual exposure can provide opportunities for forming experience-based associations, these opportunities did not consistently predict repertoire overlap, suggesting that learning via trial and error is insufficient to explain convergence in gestural repertoires.

Rather, the association between self-relevance cues and secondary contexts and responses is consistent with the possibility that these cues facilitate interpretation of gestures when signallers future behaviour cannot be predicted solely from the ongoing situation (Graham et al. 2020). We show that across homogenous gestures, secondary response type to the gesture was predicted by the presence of a social relationship based on unreciprocated grooming between the signaller and the recipient. This finding is consistent with

the interpretation that trial and error learning less readily explains the association between self-relevance cues and overlap, relative to inferential processes. Inferential accounts propose that recipients move beyond the observable behaviour itself and infer the unobservable goals, motivations, or intentions underlying visual bodily gestures (Bargh 1990; Koster-Hale and Saxe 2013). Hence, coordination contexts in which behavioral alignment is achieved through exposure to ongoing contextual cues rather than through inference of the goals, may be associated with non-overlapping visual bodily gestures rather than convergence in gestural repertoires.

This interpretation is consistent with the finding that coordination contexts such as mutual grooming can occur in the absence of gesture convergence, compatible with understanding that direct experience is not a prerequisite for gesture overlap. Instead, repertoire convergence is associated with conditions of increased uncertainty, consistent with interpretation that self-relevance cues facilitate inferential processing and alignment of communicative goals beyond what can be directly retrieved by ongoing contextual cues alone (Bouton 1994; Graham et al. 2020). Previous work combining visual bodily and manual visual gestures reported greater homogeneity in contexts such as mutual grooming, allowing us to propose that association between self-relevance cues and repertoire overlap in this context reflects an enhanced ability to infer the goal relative to visual bodily gestures unaccompanied by these behaviors (Roberts and Roberts 2018).

In support of this inferential account, previous work has shown that overlap in chimpanzee visual gestural repertoires is positively associated with features of intentional gestures such as response waiting, persistence, audience checking (Roberts and Roberts 2022), as well as accompanying manual visual gestures and vocalisations (AIR, unpubl. data). These associations are not easily explained by mechanisms of enhanced exposure to gesture–outcome contingencies because their effects remain evident even when opportunities for exposure, such as proximity and visual attention fail to predict convergence. Instead, these associations are more consistent with accounts proposing that self-relevance cues support integration of social information and inference of communicative goals under conditions of uncertainty.

This interpretation is reinforced by the finding that dyads producing gestures in secondary contexts at higher rates showed greater repertoire overlap, whereas gesture production in primary contexts was unrelated to overlap. If communicative convergence emerged primarily through reinforcement of primary gesture–outcome associations, overlap should have been most strongly associated with gestures occurring in primary contexts. Instead, convergence was specifically associated with secondary contexts where inferential processing should be most necessary. Moreover, individuals who produced more gestures in secondary contexts also produced more gestures in primary contexts, suggesting that secondary-context signalling reflects communicative flexibility needed beyond primary signalling contexts to achieve overlap in gestures. This pattern is consistent with the proposal that self-relevance cues facilitate inference of communicative goals in situations where direct experience is insufficient (Roberts et al. 2022a), potentially allowing individuals to progressively align communication and goal representations with social partners, thereby promoting repertoire overlap.

It could be argued that self-relevance cues function at a basic perceptual level to facilitate interaction between one chimpanzee and another. For example, when the density of the social audience is high, use of self-relevance cues could enhance detection of a visual bodily gesture in crowded conditions (Conty and Grèzes 2012). Consequently, self-relevance cues should be more common under conditions of high socio-ecological complexity, where the interference from social and ecological conditions is high. While several autonomic and motor responses can be potentially triggered by self-relevance cues through subcortical reflexes that are largely automatic and unconscious (Sokolov 1990), more recent work indicates that orienting reflex involves a complex interaction between cortical systems specialized for higher order cognitive operations, rather than just stimulus-driven reorienting (Corbetta et al. 2008; Patel et al. 2019). These operations include the post-perceptual updating of internal models of environmental context (Geng and Vossel 2013), conscious awareness of novel visual stimuli (Beauchamp et al. 2012; Webb et al. 2016a) and theory of mind operations (Koster-Hale and Saxe 2013). These newer formulations elevate the complexity of the perceptual mechanism underlying stimulus driven orienting reflex to a mechanism related to a network coordination and prediction (Patel et al. 2019). These prediction-error theories are based on the observation that unpredicted or unexpected stimuli will generate extra neural activity to update the internal model when compared with predicted or expected stimuli. Importantly, subjective awareness—operationalized as 'knowing' that a stimulus is being attended to—can be distinguished from simple reorienting reflex involved in dominant (primary) associations between stimuli and outcomes (Wilterson et al. 2021). Subjective awareness enables flexible allocation of attention to weaker, secondary representations, a cognitive capacity required for inferring social goals under uncertainty (Webb and Graziano 2015; Webb et al. 2016b).

Empirical work further suggests that socio-ecological complexity and increased cognitive load can suppress the neural systems associated with subjective awareness. For

instance, increasing visual short-term memory load reduces activity in the temporoparietal junction (TPJ) and impairs conscious detection of unexpected visual stimuli, producing inattentional blindness by disrupting the ventral frontoparietal attention network (Todd et al. 2005; Corbetta et al. 2008). Under these conditions, self-relevance cues are expected to buffer against interference by enhancing subjective awareness and memory (Sui and Humphreys 2015). For example we are more likely to focus on and remember someone who is looking directly at us (Conty and Grèzes 2012) or calling our name (Kampe et al. 2003).

In line with this concept, overlap in the repertoire of visual bodily signals between chimpanzee dyads was greater in conditions of lower social and ecological distraction. Although repertoire overlap was generally greater under conditions of reduced social and ecological distraction, inferential accounts predict that the mechanisms generating overlap should be most evident in socially and ecologically uncertain situations (Roberts and Roberts 2022; Xing et al. 2022; Roberts 2025). Consistent with this prediction, heterogeneous gestures were more likely in socially complex contexts and were accompanied by self-relevance cues, while these self-relevance cues predicted secondary contexts of heterogeneous gestures, and gestures produced in secondary contexts subsequently predicted greater repertoire overlap. Collectively, these patterns are more consistent with accounts proposing that self-relevance cues are associated with gestural overlap through inference of communicative goals across a wide range of social contexts than with accounts based on trial-and-error learning of gesture–outcome contingencies (Plooij 1979; Halina et al. 2013; Pika and Fröhlich 2019).

The observed patterns are compatible with the Communicative Roots of Complex Sociality and Cognition Hypothesis, which proposes that inference of the goals underlying social interactions drives both convergence and divergence in primate gestural repertoires (Roberts and Roberts 2020, 2022). In large and dynamic social networks, maintaining social cohesion may depend on individuals' ability to acquire, and flexibly deploy communicative signals with less frequently encountered or weakly bonded individuals. Within this framework, communicative complexity is proposed to increase because behavioural outcomes cannot be reliably evoked by direct interaction history, placing greater demands on mechanisms that support flexible social understanding. Inferential processes that enable individuals to track the goals underlying social interactions may therefore account for both convergence (homogeneity) and divergence (heterogeneity) in open-ended gestural repertoires, whereas experience-based accounts instead rely on repeated exposure to gesture–outcome contingencies without representing communicative goals.

Taken together, our findings are more consistent with theoretical interpretation that overlap in chimpanzee visual bodily gesture repertoires is shaped by inference of goals under uncertainty, with self-relevance cues facilitating flexible interpretation of communicative intent across diverse social partners. This has broader implications for understanding the evolution of social complexity, as it is consistent with the suggestion that communicative systems may be specifically adapted to manage time and cognitive demands in fission–fusion societies where individuals must maintain differentiated social relationships (Aiello and Dunbar 1993; Freeberg et al. 2012). Comparative research has highlighted several unresolved questions, including why gestural repertoires can be simultaneously highly variable within groups yet overlap between groups (Fröhlich et al. 2025b), why increasing social complexity is associated with greater communicative diversification (McComb and Semple 2005), and why vocal phenotypes vary with population density in orangutans (Lameira et al. 2022). Our findings offer one possible interpretation of these broader patterns by suggesting that communicative behaviour in socially complex species may be shaped not only by trial-and-error learning of signal–outcome contingencies but also by inferential processes that support flexible communication under uncertainty.

In this context, cultural complexity is often defined as the number and diversity of behavioural elements characterising a social group (Chick 1997). Socially complex societies are therefore expected to exhibit increasingly differentiated cultural repertoires. At the same time, maintaining shared behavioural repertoires may facilitate coordination among social groups. One possible interpretation is that similarity and differentiation in gestural behaviour represent opposite ends of a continuum shaped by inferential processes. Under this interpretation, inference of shared social goals would give rise and maintain convergence, whereas divergence may emerge where there are differing social goals.

A similar balance between differentiation and overlap has been proposed for other domains of primate behavior such as food culture. For instance, bonobos consume fruit, meat and foliage, and these food types vary considerably in value. The degree of refinement in texture, taste or fat content as well distinctiveness in quantity or quality define valuable foods. Socially consuming large quantities of valuable staple foods corresponds to group identities that are trusting of strangers, whereby individuals can form social connections they can depend on for support. On the other hand, very high quality valuable foods create or enhance exclusivity in line with these foods representing forms of social ranking (van der Veen 2003). In primate species such as bonobos meat may be more valuable than fruit, whereas fruit may be more valuable than foliage because

of differences in acquisition cost. The acquisition of meat culture may be more costly because of its heavier reliance on repeated interaction with social partners than acquisition of fruit culture. Differences in value would be demonstrated by in-group favouritism at out-group cost and divergence in group-specific preferences for meat (Nettle and Dunbar 1997). In contrast, out-group preferences at in-group cost would result in convergence in between group preferences for fruit. In line with this, bonobo groups appear to differ in consumption of meat (Samuni et al. 2020), while showing greater overlap in consumption of fruit between the groups (Roberts and Roberts 2026).

This combination of differentiation and overlap suggests that cultural systems may maintain both similaritiesand differences in gestural communication between the groups. In case of group-specific or localized traditions, direct experience with social partnersshapes differentiation between the groups, particularly where communication is context-specific or based on unambiguous outcomes. Simultaneously, incase of population-specific traditions, similarity in communication between the groups would be based on context-inspecific, ambiguous communicationand facilitate social bonding across groups within the population. Although the present study does not directly test these evolutionary processes, the observed behavioural patterns are consistent with theoretical proposals that cultural and gestural complexity may evolve most readily in partially connected societies, where individuals cannot solely rely on repeated direct interactions for social bonding but benefit from mechanisms supporting flexible inference about others' communicative goals (Derex and Mesoudi 2020).

The analyses in this study were carried out on a rich dataset of gestural communication in wild chimpanzees which included variables relating to social and ecological factors, self-relevance cues and the nature of the relationship between the signaller and receiver. Whilst the GLMMs made full use of the gesture-level data, there were some limitations. The observations were derived from repeated measurements of 12 focal chimpanzees. The mixed-model structure accounted for this hierarchical sampling by including random effects for signaller identity and signaller–recipient dyads, thereby reducing non-independence among observations. Nevertheless, the effective sample size remains limited, meaning that statistical precision is lower than might be inferred from the total number of gestures alone. Consequently, the reported effect sizes and significance tests should be interpreted cautiously.

A further limitation concerns the complexity of several regression models. Some of the GLMMs had a relatively large number of predictor variables, which can lead to issues with model stability and interpretability. Further, for some of the GLMMs, the VIFs obtained from corresponding linear regression analysis indicated potential issues with multicollinearity, with VIF values of between 5 and 10, although no VIF values were above the threshold of 10 which would indicate serious multicollinearity issues. These levels of collinearity are unlikely to invalidate the models but may increase uncertainty around individual regression coefficients, and therefore emphasis should be placed on the consistency of patterns across complementary analyses rather than on any single parameter estimate.

Further, it is important to distinguish the levels at which the present analyses address communicative processes. Associations between self-relevance cues and contextual variables were evaluated at the level of individual gestural events, whereas convergence and divergence in gestural repertoires were assessed at the dyadic level as cumulative patterns emerging across many communicative interactions. Consequently, the analyses cannot determine whether any single gesture event directly produces convergence or divergence. Instead, they identify statistical associations between event-level communicative features and longer-term dyadic patterns of repertoire overlap that are consistent with either inferential or experience-based accounts. While mental and attentional states can only be inferred in non-verbal animals by assessing external behaviour, we also acknowledge that, given the observational design, the results should be interpreted as associative rather than causal. Thus, the findings should be treated with caution until they are replicated in other wild chimpanzee populations, or extended to other great ape species.

In conclusion, the present findings are consistent with the prediction that self-relevance cues are associated with overlap in chimpanzee visual bodily gesture repertoires by supporting the inference of communicative goals under conditions of social and ecological uncertainty (Roberts 2025). Rather than functioning primarily to increase exposure to gesture–outcome contingencies, these self-relevance cues are proposed to facilitate flexible alignment of communication across varying social partners and contexts, particularly when direct experience is insufficient (Spocter et al. 2010). These results provide a parsimonious account of the communicative convergence and divergence in dynamic fission–fusion society of chimpanzees and highlight self-relevance cues as a potential component of flexibility in primate gestural communication.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00265-026-03796-4.

Acknowledgements We thank the Royal Zoological Society of Scotland for providing core funding to the Budongo Conservation Field Station, Uganda, as well as the Ugandan Wildlife Authority and the Uganda National Council for Science and Technology for permission to carry out this research. We would like to thank two anonymous

reviewers for their time, expertise, and thoughtful critique of this study.

Author contributions Anna Roberts conceptualized the paper, analysed the data and wrote the paper Sam Roberts analysed data and revised final version of the manuscript Piotr Sorokowski revised and corrected final version of the manuscript.

Funding AIR was funded by the Wenner-Gren Foundation for Anthropological Research, Inc. (Hunt Postdoctoral Fellowship Gr. 10567). Further, this work has been funded by National Science Centre, Poland, Grant Number: UMO-2018/31/ D/NZ8/01144 ('Understanding origins of social brains and communication in wild primates') at Institute of Human Biology and Evolution, Faculty of Biology, Adam Mickiewicz University in Poznan, 61614, Poznan, Poland.

Data availability All raw data can be found in ESM 11 - 16.

Declarations

Ethics Approval The programme of research was approved by the Animal Research Ethics Board of Stirling University, United Kingdom. In addition, all applicable international and national guidelines for the use of animals were followed.

Competing interests The authors declare no competing interests.

References

Aiello LC, Dunbar RIM (1993) Neocortex size, group size and the evolution of language. Curr Anthropol 34:184–193

Amici F, Ersson-Lembeck M, Holodynski M, Liebal K (2025) Initiation of affiliative interactions and visual engagement during mother-infant interactions in humans, great apes and small apes. Int J Primatol 46:1077–1102

Aureli F, Preston SD, de Waal FBM (1999) Heart rate responses to social interactions in free-moving rhesus macaques (Macaca mulatta): a pilot study. J Comp Psychol 113:59–65

Bakeman R, Gottman JM (1997) Observing Interaction: An Introduction to Sequential Analysis. Cambridge University Press, New York

Bard KA (1992) Intentional behavior and intentional communication in young free-ranging orangutans. Child Dev 63:1186–1197

Bargh JA (1990) Goal and intent: Goal-directed thought and behavior are often unintentional. Psychol Inq 1:248–251

Baron-Cohen S (1994) How to build a baby that can read minds: Cognitive mechanisms in mindreading. Cah Psychol Cogn 13:513–552

Barrault C, Soldati A, Hobaiter C, Mugisha S, De Moor D, Zuberbühler K, Dezecache G (2022) Thermal imaging reveals social monitoring during social feeding in wild chimpanzees. Phil Trans R Soc B 377:20210302

Bateson M, Martin P (2021) Measuring behaviour: an introductory guide. Cambridge University Press, Cambridge

Beauchamp MS, Sun P, Baum SH, Tolias AS, Yoshor D (2012) Electrocorticography links human temporoparietal junction to visual perception. Nat Neurosci 15:957–959

Behringer V, Deschner T, Möstl E, Selzer D, Hohmann G (2012) Stress affects salivary alpha-amylase activity in bonobos. Physiol Behav 105:476–482

Bethell E, Holmes A, MacLarnon A, Semple S (2012) Cognitive bias in a non-human primate: husbandry procedures influence cognitive indicators of psychological well-being in captive rhesus macaques. Anim Welf 21:185–195

Borgatti SP, Everett MG, Freeman LC (2014) Ucinet. In: Alhajj R, Rokne J (eds) Encyclopedia of Social Network Analysis and Mining. Springer-, New York, pp 2261–2267

Bouton ME (1994) Context, ambiguity, and classical conditioning. Curr Dir Psychol Sci 3:49–53

Bouton ME (2021) Context, attention, and the switch between habit and goal-direction in behavior. Learn Behav 49:349–362

Carrasco M, Ling S, Read S (2004) Attention alters appearance. Nat Neurosci 7:308–313

Cartmill EA (2023) Overcoming bias in the comparison of human language and animal communication. P Natl Acad Sci USA 120:e2218799120

Chick G (1997) Cultural complexity: The concept and its measurement. Cross-Cult Res 31:275–307

Conty L, Grèzes J (2012) Look at me, I'll remember you: the perception of self-relevant social cues enhances memory and right hippocampal activity. Hum Brain Map 33:2428–2440

Corbetta M, Patel G, Shulman GL (2008) The reorienting system of the human brain: from environment to theory of mind. Neuron 58:306–324

Coussi-Korbel S, Fragaszy DM (1995) On the relation between social dynamics and social learning. Anim Behav 50:1441–1453

Croft DP, James R, Krause J (2010) Exploring Animal Social Networks. Princeton University Press, Princeton, NJ

Davis FC, Neta M, Kim MJ, Moran JM, Whalen PJ (2016) Interpreting ambiguous social cues in unpredictable contexts. Social Cogn Aff Neurosci 11:775–782

Derex M, Mesoudi A (2020) Cumulative cultural evolution within evolving population structures. Trends Cogn Sci 24:654–667

Dunbar RIM (1998) The social brain hypothesis. Evol Anthropol 6:178–190

Dunbar R (2025) Structural and cognitive mechanisms of group cohesion in primates. Behav Brain Sci 48:e162

Freeberg TM, Dunbar RIM, Ord TJ (2012) Social complexity as a proximate and ultimate factor in communicative complexity. Phil Trans R Soc B 367:1785–1801

Fröhlich M, Aychet J, Clark PR, Crockford C, Dezecache G, Rebout N, van Schaik CP, Graham KE (2025a) Assessing flexibility in meaning and context in non-human communication. Biol Rev 100:2471–2481

Fröhlich M, Boeckx C, Tennie C (2025b) The role of exploration and exploitation in primate communication. Pro R Soc B 292:20241665

Geng JJ, Vossel S (2013) Re-evaluating the role of TPJ in attentional control: contextual updating? Neurosci Biobehav Rev 37:2608–2620

Genty E, Breuer T, Hobaiter C, Byrne RW (2009) Gestural communication of the gorilla (*Gorilla gorilla*): Repertoire, intentionality and possible origins. Anim Cogn 12:527–546

Gomez JC (2026) Ostensive communication in great apes: The evolution of Gricean intent. Mind Lang. https://doi.org/10.1111/mila.70036. published online

Graham KE, Furuichi T, Byrne RW (2020) Context, not sequence order, affects the meaning of bonobo (*Pan paniscus*) gestures. Gesture 19:335–364

Halina M, Rossano F, Tomasello M (2013) The ontogenetic ritualization of bonobo gestures. Anim Cogn 16:653–666

Hewes GW (1973) Primate communication and the gestural origin of language. Curr Anthropol 14:5–24

Hobaiter K, Byrne R (2011) The gestural repertoire of the wild chimpanzee. Anim Cogn 14:745–767

Hopkins WD, Russel JL, Cantalupo C (2007) Neuroanatomical correlates of handedness for tool use in chimpanzees (*Pan Troglodytes*): Implication for theories on the evolution of language. Psychol Sci 18:971–977

Kampe KK, Frith CD, Frith U (2003) Hey John: signals conveying communicative intention toward the self activate brain regions associated with mentalizing, regardless of modality. J Neurosci 23:5258–5263

Koster-Hale J, Saxe R (2013) Theory of mind: a neural prediction problem. Neuron 79:836–848

Kudo H, Dunbar RIM (2001) Neocortex size and social network size in primates. Anim Behav 62:711–722

Kutner M, Nachtsheim C, Neter J, Li W (2005) Applied linear statistical models. McGraw-Hill, New York

Laland KN, Janik VM (2006) The animal cultures debate. Trends Ecol Evol 21:542–547

Lameira AR, Santamaría-Bonfil G, Galeone D, Gamba M, Hardus ME, Knott CD, Morrogh-Bernard H, Nowak MG, Campbell-Smith G, Wich SA (2022) Sociality predicts orangutan vocal phenotype. Nat Ecol Evol 6:644–652

Landis JR, Koch GG (1977) An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 33:363–374

Leavens DA, Russell JL, Hopkins WD (2010) Multimodal communication by captive chimpanzees (*Pan troglodytes*). Anim Cogn 13:33–40

Lewis LS, Kano F, Stevens JM, DuBois JG, Call J, Krupenye C (2021) Bonobos and chimpanzees preferentially attend to familiar members of the dominant sex. Anim Behav 177:193–206

Liebal K, Call J, Tomasello M (2004) Use of gesture sequences in chimpanzees. Am J Primatol 64:377–396

Lyn H, Russell JL, Leavens DA, Bard KA, Boysen ST, Schaeffer JA, Hopkins WD (2014) Apes communicate about absent and displaced objects: methodology matters. Anim Cogn 17:85–94

McComb K, Semple S (2005) Coevolution of vocal communication and sociality in primates. Biol Lett 1:381–385

Mine JG, Wilke C, Zulberti C, Behjati M, Bosshard AB, Stoll S, Machanda ZP, Manser A, Slocombe KE, Townsend SW (2024) Vocal-visual combinations in wild chimpanzees. Behav Ecol Sociobiol 78:108

Mitani JC, Watts DP, Pepper JW, Merriwether DA (2002) Demographic and social constraints on male chimpanzee behaviour. Anim Behav 64:727–737

Nellissen L, Fuh T, Masi S, Zuberbühler K (2026) Multimodal coordination of group departures in western gorillas. Anim Behav 235:123551

Nettle D, Dunbar RIM (1997) Social markers and the evolution of reciprocal exchange. Curr Anthropol 38:93–99

Noudoost B, Moore T (2011) The role of neuromodulators in selective attention. Trends Cogn Sci 15:585–591

Patel GH, Sestieri C, Corbetta M (2019) The evolution of the temporoparietal junction and posterior superior temporal sulcus. Cortex 118:38–50

Payne JD, Jackson ED, Hoscheidt S, Ryan L, Jacobs WJ, Nadel L (2007) Stress administered prior to encoding impairs neutral but enhances emotional long-term episodic memories. Learn Mem 14:861–868

Pedruzzi L, Galotti A, Francesconi M, Quartesan A, Gamessa SA, Argaw HT, Meletiadis A, Acutis PL, Riina MV, Serra V (2026) Relationship value, emotion, and uncertainty shape conflict–resolution communication in a wild primate. Ann NY Acad Sci 1562:e70352

Pika S, Fröhlich M (2019) Gestural acquisition in great apes: the social negotiation hypothesis. Anim Cogn 22:551–565

Pika S, Tomasello M (2002) Gestural communication in gorillas (*Gorilla g. gorilla*) and bonobos (*Pan paniscus*) in captivity: A comparison. Adv Ethol 37:65

Plooij FX (1978) Some basic traits of language in wild chimpanzees. In: Lock A (ed) Action, gesture and symbol: The emergence of language. Academic, London, pp 111–131

Plooij FX (1979) How wild chimpanzee babies trigger the onset of mother-infant play - and what the mother makes of it. In: Bullowa M (ed) Before speech: the beginning of interpersonal communication. Cambridge University Press, New York, pp 223–243

Roberts AI (2025) Updating social knowledge under stress: communicative complexity and the evolution of primate social bonding. Neurosci Biobehav Rev 179:106432

Roberts AI, Roberts SGB (2015) Gestural communication and mating tactics in wild chimpanzees. PLoS ONE 10:e0139683

Roberts AI, Roberts SGB (2016a) Wild chimpanzees modify modality of gestures according to the strength of social bonds and personal network size. Sci Rep 6:33864

Roberts SGB, Roberts AI (2016b) Social brain hypothesis, vocal and gesture networks of wild chimpanzees. Front Psychol 7:1756

Roberts AI, Roberts SGB (2017) Convergence and divergence in gesture repertoires as an adaptive mechanism for social bonding in primates. R Soc Open Sci 4:170181

Roberts SGB, Roberts AI (2018) Visual attention, indicative gestures, and calls accompanying gestural communication are associated with sociality in wild chimpanzees (*Pan troglodyres schweinfurthii*). J Comp Psychol 133:56–71

Roberts AI, Roberts SGB (2019) Persistence in gestural communication predicts sociality in wild chimpanzees. Anim Cogn 22:605–618

Roberts AI, Roberts SGB (2020) Communicative roots of complex sociality and cognition. Biol Rev 95:51–73

Roberts AI, Roberts SGB (2022) Intentional gesturing increases social complexity by allowing recipient's understanding of intentions when it is inhibited by stress. Phil Trans R Soc B 377:20210305

Roberts AI, Roberts SGB (2026) Food culture as a mechanism of social bonding and social identity in primates. Front Hum Neurosci 20:1788917

Roberts AI, Vick S-J, Buchanan-Smith H (2012a) Usage and comprehension of manual gestures in wild chimpanzees. Anim Behav 84:459–470

Roberts AI, Vick S-J, Roberts SGB, Buchanan-Smith HM, Zuberbühler K (2012b) A structure-based repertoire of manual gestures in wild chimpanzees: Statistical analyses of a graded communication system. Evol Hum Behav 33:578–589

Roberts AI, Roberts SGB, Vick S-J (2014) The repertoire and intentionality of gestural communication in wild chimpanzees. Anim Cogn 17:317–336

Roberts SGB, Dunbar RIM, Roberts AI (2022a) Communicative roots of complex sociality and cognition: Neuropsychological mechanisms underpinning the processing of social information. Phil Trans R Soc B 377:20210295

Roberts SGB, Dunbar RIM, Roberts AI (2022b) Communicative roots of complex sociality and cognition: neuropsychological

mechanisms underpinning the processing of social information. Phil Trans R Soc B 377:20210295
Roelofs S, Boleij H, Nordquist RE, van der Staay FJ (2016) Making decisions under ambiguity: judgment bias tasks for assessing emotional state in animals. Front Behav Neurosci 10:119
Samuni L, Wegdell F, Surbeck M (2020) Behavioural diversity of bonobo prey preference as a potential cultural trait. Elife 9:e59191
Schleihauf H, Sanford EM, Thompson BD, Zhang S, Rukundo J, Call J, Herrmann E, Engelmann JM (2025) Chimpanzees rationally revise their beliefs. Science 390:521–526
Shultz S, Dunbar RIM (2022) Socioecological complexity in primate groups and its cognitive correlates. Phil Trans R Soc B 377:20210296
Sokolov E (1990) The orienting response, and future directions of its development. Pavlov J Biol Sci 25:142–150
Sperber D, Wilson D (2026) Rethinking ostensive communication in an evolutionary, comparative, and developmental perspective. Psychol Rev 133:315
Spocter MA, Hopkins WD, Garrison AR, Bauernfeind AL, Stimpson CD, Hof PR, Sherwood CC (2010) Wernicke's area homologue in chimpanzees (*Pan troglodytes*) and its relation to the appearance of modern human language. Proc R Soc B 277:2165
Sui J, Humphreys GW (2015) The integrative self: How self-reference integrates perception and memory. Trends Cogn Sci 19:719–728
Taglialatela JP, Russell JL, Pope SM, Morton T, Bogart S, Reamer LA, Schapiro SJ, Hopkins WD (2015) Multimodal communication in chimpanzees. Am J Primatol 77:1143–1148
Todd JJ, Fougnie D, Marois R (2005) Visual short-term memory load suppresses temporo-parietal junction activity and induces inattentional blindness. Psychol Sci 16:965–972
Tomasello M (1999) The Cultural Origins of Human Cognition. Harvard University Press, Cambridge, MA
Tomasello M, Call J, Nagell K, Olguin R, Carpenter M (1994) The learning and use of gestural signals by young chimpanzees: A trans-generational study. Primates 37:137–154
Tomasello M, Call J, Warren J, Frost T, Carpenter M, Nagell K (1997) The ontogeny of chimpanzee gestural signals: A comparison across groups and generations. Evol Communication 1:223–253
Tomasello M, Carpenter M, Call J, Behne T, Moll H (2005) Understanding and sharing intentions: the origins of cultural cognition. Behav Brain Sci 28:675–735
Tomasello R, Kim C, Dreyer FR, Grisoni L, Pulvermüller F (2019) Neurophysiological evidence for rapid processing of verbal and gestural information in understanding communicative actions. Sci Rep 9:16285
van Boekholt B, Wilkinson R, Pika S (2024) Bodies at play: the role of intercorporeality and bodily affordances in coordinating social play in chimpanzees in the wild. Front Psychol 14:1206497
van der Veen M (2003) When is food a luxury? World Archaeol 34:405–427
van Schaik CP, Isler K, Burkart JM (2012) Explaining brain size variation: from social to cultural brain. Trends Cogn Sci 16:277–284
Vick S-J, Waller BM, Parr LA, Smith Pasqualini MC, Bard KA (2007) A cross-species comparison of facial morphology and movement in humans and chimpanzees using the facial action coding system (FACS). J Nonverb Behav 31:1–20
Watson CF, Buchanan-Smith HM, Caldwell CA (2014) Call playback artificially generates a temporary cultural style of high affiliation in marmosets. Anim Behav 93:163–171
Webb TW, Graziano MS (2015) The attention schema theory: a mechanistic account of subjective awareness. Front Psychol 6:500
Webb TW, Igelström KM, Schurger A, Graziano MS (2016a) Cortical networks involved in visual awareness independent of visual attention. P Natl Acad Sci USA 113:13923–13928
Webb TW, Kean HH, Graziano MS (2016b) Effects of awareness on the control of attention. J Cogn Neurosci 28:842–851
Whiten A (2021) The burgeoning reach of animal culture. Science 372:eabe6514
Wilke C, Kavanagh E, Donnellan E, Waller BM, Machanda ZP, Slocombe KE (2017) Production of and responses to unimodal and multimodal signals in wild chimpanzees, *Pan troglodytes schweinfurthii*. Anim Behav 123:305–316
Wilterson AI, Nastase SA, Bio BJ, Guterstam A, Graziano MS (2021) Attention, awareness, and the right temporoparietal junction. P Natl Acad Sci USA 118:e2026099118
Xing W, Zhang S, Wang Z, Jiang D, Han S, Luo Y (2022) Self-awareness protects working memory in people under chronic stress: An ERP study. Front Psychol 13:1003719
Zajonc RB, Sales SM (1966) Social facilitation of dominant and subordinate responses. J Exp Soc Psychol 2:160–168

Authors and Affiliations

Anna Ilona Roberts[1,2] · Piotr Sorokowski[1] · Sam George Bradley Roberts[3]

✉ Anna Ilona Roberts
anna.roberts@uwr.edu.pl

[1] Institute of Psychology, University of Wrocław, Dawida 1, Wrocław 50-527, Poland

[2] Bonobo Foundation, 2 Clare Avenue, Chester CH2 3HS, UK

[3] School of Psychology, John Moores University, Byrom Street, Liverpool L3 3AF, UK