Title

Latent meaning representations in great-ape gestural communication

Permalink

https://escholarship.org/uc/item/56c0m5t0

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Franke, Michael

Bohn, Manuel

Fröhlich, Marlen

Publication Date

2024

Peer reviewed

Latent meaning representations in great-ape gestural communication

Michael Franke (michael.franke@uni-tuebingen.de)
Department of Linguistics, University of Tübingen

Manuel Bohn (manuel.bohn@leuphana.de)
Institute of Psychology in Education, Leuphana University Lüneburg

Marlen Fröhlich (marlen.froehlich@uni-tuebingen.de)
Institute for Archaeological Sciences, University of Tübingen

Abstract

Studies of meaning in human and primate communication face, in principle, similar methodological problems. In both cases, meaning is not observable directly, but must be inferred from more indirect sources, such as directly observable behavior. Recent work in probabilistic cognitive modeling of language use has therefore developed methods of inferring latent semantic meaning through the lens of a probabilistic model of language use. In this paper, we explore how to adapt such an approach for insightful investigations of primate communication. Towards this end, we develop a suitable probabilistic model of processes that generate communicative behavior by making use of functionally specified latent meaning representations. As a proof of concept, we apply this model to a rich, annotated data set of orangutan communicative dyadic interaction and conclude that explicit probabilistic modeling can provide additional insights for the study of animal communication pertaining to the context-dependent nature of signals and the gradual evolution of human communication systems.

Keywords: primate communication; evolution of language; pragmatics; probabilistic modeling

Introduction

Language must have evolved in the primate order, but how and why this happened remains one of the big scientific questions of our time. A core feature of language is its flexibility, allowing humans to effortlessly vary and switch between lexical items under different social and interactional conditions. The degree to which nonhuman species are able to adjust their communicative signals and their responses to them according to the immediate socio-ecological condition has critical implications for the evolutionary trajectory of pragmatics, that is, the role of context in shaping the meaning of utterances (Scott-Phillips, 2010; Wheeler & Fischer, 2012). In this paper, we want to build common ground between the study of pragmatics in human and nonhuman communication by explicating a computational framework that specifies a basic architecture of context-sensitive inferential communication.

Determining the meaning of signals has been a central challenge in comparative research, complicated by the common conflation between context and function (Fröhlich et al., under review) but also by the discrepancy in methodological approaches between research fields (Liebal & Oña, 2018). Vocal communication studies often use playback experiments to test hypotheses, and to date mainly focused on the representation and functional reference of signals (Wheeler & Fischer, 2012). In contrast, gestural communication researchers conduct observational studies of behavior around communication events to infer intentional signal use (Liebal et al., 2014). For a long time, gestures have been characterized as extraordinarily flexible (Corballis, 2002), but evidence for vocal flexibility in several primate taxa is now increasing (Taylor et al., 2022). The rise of "multimodalism" in comparative research on nonhuman communication systems (Fröhlich et al., 2021a) requires adopting an integrative definition of meaning applicable across communicative modalities.

There seems to be an insurmountable gulf between nonhuman communication, where signal and function are biologically fixed (Smith, 1977), and human language, which relies on higher-order mental representations (Grice, 1957). A bridge across this gap may be provided by communication systems in which signals are used intentionally, that is signallers produce signals to change the behaviour of recipients, evidencing so-called "first-order intentionality" (Dennett, 1983). Studying gesture meanings in wild chimpanzees, Hobaiter & Byrne (2014) noted the presence of apparently satisfactory outcomes (ASO) after which signalling stops, thus reflecting plausible social goals. These and other studies focusing on audience checking, sensitivity to visual attention, and persistence (Leavens et al., 2005; Fröhlich et al., 2019) showed that the use of signals in nonhuman primates can be intentional and voluntary. The presence of an intended outcome implies that the link between internal state and behavioural outcome is potentially uncoupled and that the signal is primarily linked to the goal. The *selected effect* ("function") becomes the *intended effect*, and ASOs thus reflect the "Resultant Meaning Facet" sensu Amphaeris et al. (2023).

Recent years have seen growing evidence for context dependence in the use and interpretation of signals in nonhuman communication, spanning multiple primate taxa and communicative modalities. For instance, comparative research on great ape gestures has shown that often there is no one-to-one correspondence between signal and function in primates (Liebal et al., 2014), as commonly assumed for the code-based model in animal communication (Cartmill, 2023), and that the interpretation of signals is typically 'enriched' using cues external to the signal itself. Hence, nonhuman primates have been considered suitable models for this emerging area of research, supporting pragmatics-first approaches of language evolution (Scott-Phillips, 2010; Moore, 2016).

The study of pragmatics in nonhuman animals has largely

been isolated from research on this topic in human adults and children. Even though key terms like "meaning" or "intention" are used with reference to foundational work in linguistic pragmatics (Grice, 1957), there is little methodological and theoretical overlap between the fields. As a first step in this direction, Bohn et al. (2022) recently introduced a computational framework to bridge that gap. Building on the work on intentional communication in primates they took inspiration from one of the most influential and productive frameworks in linguistic pragmatics – the Rational Speech Act (RSA) framework (Frank & Goodman, 2012; Franke & Jäger, 2016; Scontras et al., 2021) – and suggested that communicative interactions between chimpanzees could be understood as a social inference process in which receivers (listeners) rationally integrate multiple information sources (gestures, facial expressions, social context). Yet, despite being conceptually interesting, the work of Bohn et al. (2022) remains superficial: while focusing on the high-level integration of different (multi-modal) information sources, they fail to specify the processes underlying each source. For example, the way in which a signal relates to, e.g., a sender's intention is not spelled out and instead simply represented by a stipulated number denoting associative strength.

Our goal in this paper is to close the theoretical and methodological gap dividing the study of human and non-human pragmatics. As a first step, we present a computational model that assigns a functional role to latent signal meanings in the communication process. To overcome the technical limitations of previous modeling work by Bohn et al. (2022), where meaning associations were stipulated and fixed, we draw on related work from linguistics in which probabilistic models of production and interpretation behavior are formally specified and Bayesian inference is used to model dyadic learning or adaptation between interlocutors (Schuster & Degen, 2020; Hawkins et al., 2023) or to estimate plausible semantic values from the observed data (Schöller & Franke, 2017; van Tiel et al., 2021). Next, we use this model to analyze a large data set of orangutan (*Pongo abelii, P. pygmaeus*) gestural communication, first published in Fröhlich et al. (2021b). Finally, we use this method to investigate contextual flexibility by comparing signal meanings in different communicative settings (wild vs. captive, mother-offspring dyads vs. others). Taken together, this paper constitutes a theoretical step forward by explicitly modelling the cognitive processes involved in great ape communication as well as a methodological step forward by proposing a model that can directly be used to analyse existing data.

Probabilistic models for great-ape signaling

Our goal is to uncover likely signal-meaning associations that would explain the observed dyadic behavior in the data set of orangutan gestural communication. The problem is that, while behavior is directly observed, signal-meaning associations are not; rather, they have to be inferred from the data. This becomes possible if we identify a functional role for

signal-meaning mappings in a model of the process that (is assumed to have) generated the data. Based on such a model, we can use Bayesian inference to obtain information of latent semantic meanings of signals in great-ape interaction.

To do so, we start with a model of dyadic interactions as a *signaling game* (Lewis, 1969; Skyrms, 2010). We then define choice policies for the actors in these games, which are parameterized on a numerically represented association strength between signals and meanings, referred to here as a lexicon (matrix) L. We then use the data from all dyads, sender and receiver together, to infer values of association strength L that are likely to have generated the observed data.

The stage game. A signaling game has two players, a sender S and a receiver R. It consists of a set T of states (mnemonic for *types* of the speaker), a set M of messages or signals, and a set A of response actions. In one round of the game, the sender knows the actual world state (drawn from a prior distribution $P_T \in \Delta(T)$ over states). The sender selects a message $m \in M$ conditional on the observed state $t \in T$, by sampling from the sender's policy $P_S(\cdot \mid t) \in \Delta(M)$. The receiver does not know the state, but observes the sender's choice of message, and so responds by selecting an action from the receiver's policy $P_R(\cdot \mid m) \in \Delta(A)$. For each round of play, given by a triple $\langle t, m, a \rangle$, both sender and receiver obtain a numerical utility $U_{S,R}(t, m, a) \in \mathbb{R}$.

For our analyses, we instantiate a signaling game as follows and call it the **stage game**. The set of states T consists of seven *need states*, which correspond to the seven ASOs identified in the data set. The set of signals M contains 42 different signals, as annotated in the data set. The set of receiver actions A consists of the seven types of actions which correspond directly to the need states T (thus leading to an 'ASO' classification in the data set), conjoined with all the other types of response actions distinguished during the annotation process (i.e., 'none', 'other - agonistic', 'other - move away', 'responds w/ signal', 'tolerates', 'visual attention').

Solution concepts & estimated functional policies. Standard applications of signaling games examine equilibria, evolutionary stable states or other kinds of game-theoretic solution concepts (e.g. Cho & Kreps, 1987; Wärneryd, 1993). This approach is entirely behavioral and simply looks at which policies satisfy the criteria of the solution concept. In contrast, we are here interested in how agents could internally process information and implement a given policy, thus taking a more cognitive approach to specifying behavior (Brochhagen et al., 2018). Concretely, we will spell out a minimal **lexical-association model**, which assumes that the choice policies for sender and receiver are conditioned on a latent, shared lexicon L, which is estimated from the observed data. In other words, rather than taking the fixed signaling game (without empirical data) and asking which policies best "solve" the game, we here use available empirical data to estimate which policies best explain the data, based on

a functional parameterization that is theoretically insightful about possible cognitive processes (here: meaninglike associations between states and signals) that could have generated the data. For comparison, we also fit a **behavioral baseline model**, which remains agnostic about the information processes that may feed into signaling behavior and serves as a baseline model guaranteed to give optimal fit to the data.

Behavioral baseline model. Our data consists of a series of observed plays of the stage game, i.e., triples of state, message, and action. For any given sender policy P_S and receiver policy P_R, the likelihood of observing such a triple $\langle t, m, a \rangle$ is:

$$\text{LH}(\langle t, m, a \rangle, P_S, P_R) = P_S(m \mid t) \times P_R(a \mid m).$$

The likelihood of all data D is:

$$\text{LH}(D, P_S, P_R) = \prod_{\langle t, m, a \rangle \in D} P_S(m \mid t) \times P_R(a \mid m).$$

Using Bayes' rule, we can, in principle, infer *a posteriori* credible policies:

$$P(P_S, P_R \mid D) \propto P(P_S, P_R) \times \text{LH}(D, P_S, P_R).$$

The behavioral baseline model does this by sampling policies from a sequence of unbiased Dirichlet distributions. Essentially, this model just fits, in parallel, unbiased rates for multinomial distributions for each state (sender) and message (receiver). Being unbiased in the prior, the means of posterior rate parameters approximate the corresponding maximum likelihood values. The model has a total of $7 \times 41 + 42 \times 12 = 791$ free parameters, and is maximally powerful to capture all global patterns in the data (without group distinctions, see below). For comparison, we can think of this model as two conjoined multinomial regression models, one that predicts message choice given state, and another that predicts action choice given message. Using Stan (Carpenter et al., 2016, in press), we obtain samples for the posterior distribution over multinomial rates. The mean log-likelihood for posterior samples from the policies is $-45,298$ (95% credible interval $[-45,340; -45,258]$).[1]

Lexical-association model. The lexical-association model (LAM) uses the same likelihood function, but defines policies as parameterized functions of conceptually meaningful parameters, taking inspiration from probabilsitic models of pragmatic utterance choice and interpretation for human language (e.g., Frank & Goodman, 2012). While human communicators do plausibly engage, at least sometimes, in some form of audience-design or strategic Theory-of-Mind reasoning, our starting assumption for models of great-ape signaling is that senders and receivers are *not* reasoning about each

other; they are rather modelled as what the pragmatics literature refers to as *literal agents*.[2]

The main assumption of the lexical-association model is that sender and receiver policies are conditioned on a matrix L (mnemonic for *lexicon*) of size $|T| \times |M|$, such that each entry $L_{ij} \in [0; 1]$ represents how strongly state t_i is associated with signal m_j. Crucially, the model assumes that the *same* state-message associations underlie the sender and the receiver behavior. In this way, this model goes beyond what is normally achievable, e.g., by standard regression modeling. It incorporates, and therefore allows to test, the assumption that great-ape signaling is governed by, what we may call, a **shared semantic lexicon**, that explains both production and reception at the same time. Reversely, the model pools information from both production and reception for inference, for an arguably stronger basis for inference of the meanings of signals than could be achieved by statistical results that rely on only one side of the communication channel.

Concretely, the literal sender policy is defined as:

$$P_S(m_j \mid t_i) \propto L_{ij}.$$

So, the literal sender selects a message proportionally to its association strength for the given state.

The literal receiver policy is more complex. Since lexical associations are defined between states and messages, but the receiver maps messages to actions, the policy must bring to bear information from lexical association onto their choice of action. Using ideas from decision and game theory, we assume that this mapping may happen by choosing actions based on their expected utility. Therefore, we also require a utility matrix U of size $|T| \times |M|$, such that $U_{ik} \in \mathbb{R}$ is the receiver's utility associated with action a_k when performed in state t_i. The literal receiver's policy is then given by:

$$P_R(a_k \mid m_j) = \text{Soft-Max}_\alpha(EU(a_k, m_j))$$
$$EU(a_k, m_j) = \sum_{t_i \in T} P_R(t_i \mid m_j) U_{ik}$$
$$P_R(t_i \mid m_j) \propto P_T(t_i) L_{ij}$$

In words, the policy is defined as a soft-max choice, with additional free model parameter α, based on the expected utilities (Franke & Degen, 2023). The latter are computed, in the normal way, by combining information about lexical associations with a prior distribution over states (which we here assume to be fixed, determined as the relative frequency of need states in the data set).

To fit this, we assume the following priors:

$$L_{ij} \sim \text{Beta}(0.5, 7)$$
$$U_{ik} \sim \text{Normal}(0, 0.2)$$
$$\alpha \sim \log-\mathcal{N}(2, 0.05)$$

---

[1] All model fits were obtained by Hamiltonian Monte Carlo in four chains, an initial warm-up period of 3000 samples, and recording 4000 samples for each chain. Quality of samples were checked visually, by $\hat{R}$ statistic, and for efficient sample size. Data and code are available here: https://tinyurl.com/u2uywkny

[2] If the methodology we explore here shows to be useful enough, future work should also consider more sophisticated agent models and let the data decide which ones are most adequate for data from great-ape communication.
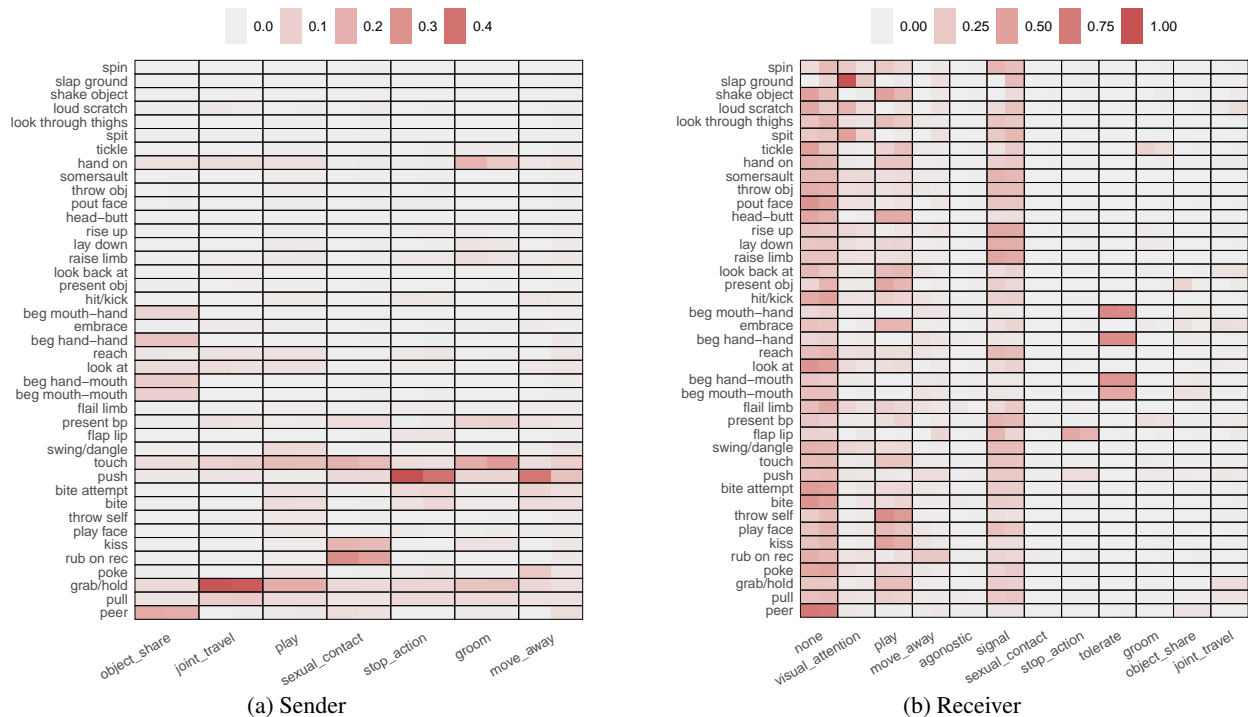
Figure 1: Visual comparison of data and posterior predictions by the LAM for sender and receiver. Each plot visualizes two matrices. Each cell shows the observed proportion of choices in the data (left), and the mean of the posterior prediction of the LAM (right). For the sender, each column is a state, so the colors in each column represent a probability distribution over messages. For the receiver, each row is a message, so the colors in each row represent a probability distribution over actions.

These priors are chosen so as to be regularizing (in principle infinitely many pairs of U and α map onto the same receiver policy), and also to implement a structural initial bias towards small lexical association. Intuitively, as a default we expect *no* association, and would require empirical data to convince us that one exists.

The resulting model has less than half the number of free parameters of the previous baseline model, namely $7 \times 42 + 7 \times 13 + 1 = 386$, which drives home the important point that the lexical-association model attempts data compression with conceptually interpretable structure: by attempting to explain production and reception with a shared association structure, we can obtain more information about the interactive nature of communication than we would based on superficial data-fitting of each side separately. Naturally, compression comes with a nominal loss in mean posterior likelihood of the data $-45,456$ ($[-45,488; -45,422]$). Indeed, the predictive performance of the baseline model is significantly better under leave-one-out model comparison (Vehtari et al., 2017): difference in expected log-probability 136.6 (SE 25.8). But notice that the LAM is able to generalize accross entire states and messages, i.e., make predictions for a held-out state when trained on all other data, while the baseline model cannot.

Figure 1 shows the observed frequencies of behavior in the data set (left part in each colored cell) and the LAM's mean posterior prediction (right part). For the most part, the LAM captures the regularities in the data reasonably well (left and right coloring in many cells are the very similar). The strong association between most messages and the action "none" on the receiver side (Fig. 1b) is striking, but merely results from the fact that orang-utans only respond to the signals of their conspecifics in about two thirds of signalling instances (Fröhlich et al., 2021a). This is consistent with reported response rates in both captive (Liebal et al., 2004) and wild (Hobaiter & Byrne, 2011) chimpanzees, and probably reflects a fundamental difference in responsiveness ("conversational" or "joint" commitment) between human and nonhuman apes. Notice that the visually striking outlier for the receiver data for signal "slap ground" is based on only three data observations for this signal in the entire data set, but that sparsity of data is not a problem for Bayesian inference. We conclude that a model assuming a shared meaning representation that feeds sender and receiver policies alike is able to reproduce coarse-grained patterns in the data.

Investigating lexical associations

A computational model that allows inference of something akin to functional "meaning representations" may help address a number of theoretically interesting research questions more directly and more clearly than other methods of data analysis. To explore these possibilities, we extend the LAM to allow for group comparisons, focussing here on two distinctions: (i) between mother-offspring and all other types of dyads, and (ii) between individuals in the wild and in captiv-
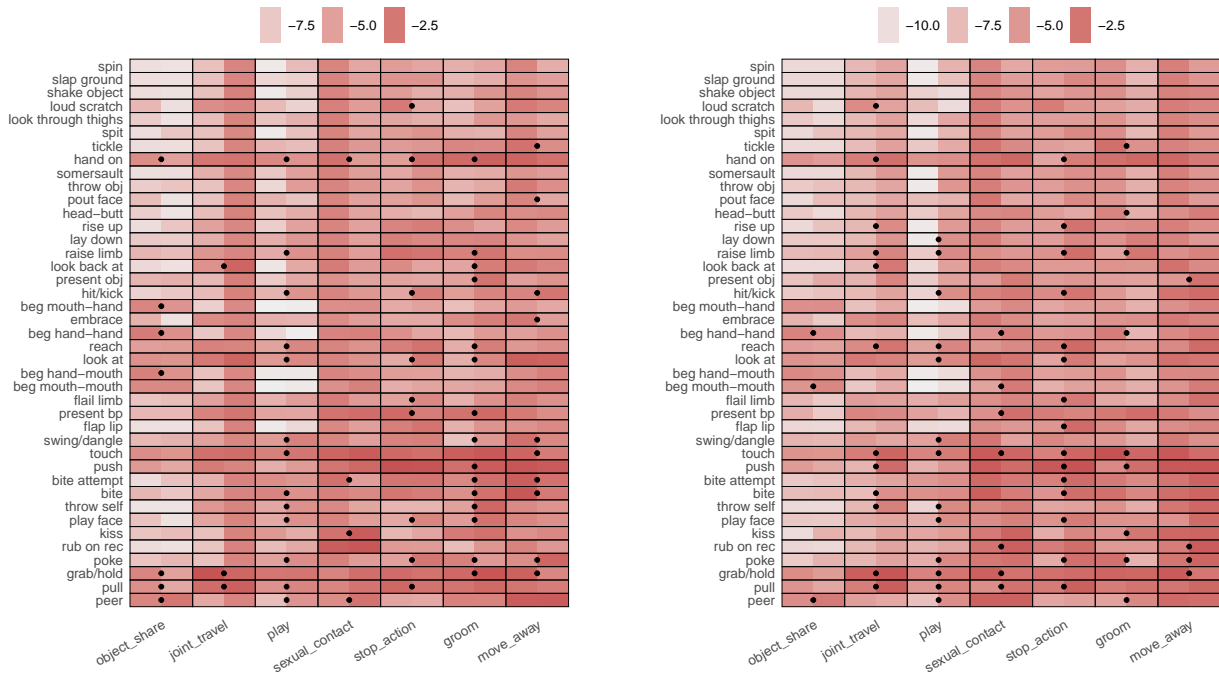
(a) 'Mother-offspring' (left) vs. 'other' (right) (b) 'Wild' (left) vs. 'captive' (right)

Figure 2: Mean posterior of lexical association (log-values for better visibility) in two-group comparisons. For each state i (x-axis) and message j (y-axis) pair, each cell shows the mean of the posterior distribution of the corresponding lexical association parameter L_{ij}, which is logged for better visual comparison of small values. The value of one group is shown on the left in each cell, the other on the right. Black dots indicate noteworthy differences between groups for any given state-message pair.

ity. For each group comparison, the LAM fits two separate lexicon matrices, one for each group (sampled independently from the same prior). Predictions for the data from each group are computed based on the lexicon for that group. The other model parameters (soft-max parameter α and the utilities) are shared between groups. We use priors over states that are calculated independently from the data from each group.

Figure 2 shows mean inferred values of lexical association, comparing different groups, and indicating noteworthy differences in lexical association. For better visibility, logarithmic values are shown. We flag a difference between inferred lexical associations as noteworthy if the posterior probability of that difference is bigger than 0.95 or smaller than 0.05, and the absolute value of the difference is bigger than 0.01. Note that a nominally large difference between the (logarithms of the) means of association values do not necessarily imply credible differences, or vice versa.

There appears to be substantial overlap in the inferred lexical associations between different groups, but also noteworthy and insightful differences. To interpret this, we need to take the functional role of lexical association values into account. A lexical value L_{ij} for state i and message j plays a double role, informing both the sender and the receiver policy. So, when interpreting inferred lexical association in Figure 2, vertical comparison (in each column) informs us about the impact on the sender's policy. Horizontal comparisons are informative about the implied receiver policy, but must be interpreted with caution, because the receiver policy also takes

state priors and utilities into account.

With this in mind, when contrasting mother-offspring with other dyads, we see that lexica differ vertically for states *play* and *grooming*. For *grooming*, mother-offspring dyads appear to have more pronounced lexical associations for some messages, in that there appear to be larger differences in vertical comparison. That is, in mother-infant dyads, senders prefer to use a smaller set of signals when communicating about this state compared to senders in other dyads. A similar pattern can be seen for *play*. This may reflect, in holistic ways (taking sender *and* receiver side into account), the intricacies of mother-infant interactions in that playing and grooming are often part of regular care-giving routines. As a consequence, receivers —taking this contextual information into account— will more readily associate *play* or *grooming* with every message and so it requires a more specific (horizontal) association of messages to successfully communicate about these states when intended. In sum, inferred latent "meaning representations" may be informative about context-specific signal use, due to their holistic functional role in explaining sender and receiver behavior.

Communication for the initiation of joint travel, on the other hand, is hardly observed beyond mother-offspring pairs. Consequently, inferred lexical association for the state *joint travel* are largely uninformative for the *other* group. This also explains why there are so few state-message pairs for which we see noteworthy differences between the two groups for that state: with few data points to draw information from,

637

inferences for the *other* group are vague, thus showing little noteworthy differences. However, the three pairs that do differ between groups may reveal an interesting pattern: physical messages like "grab/hold" and "pull" have a higher lexical association to this state in mother-offspring pairs compared to others, while the opposite is evident for the non-contact message "look back". Once again, this can be seen as a consequence of how communication is embedded in social interaction. Mother-offspring dyads spend more time in close proximity, making tactile gestures more efficient means of communication.

When comparing lexica inferred for wild vs. captive populations, we see the most differences for *joint travel*, *play* and *stop action*. In all three cases, we see more dispersed lexical associations, with more messages being associated with states for animals in captivity. One interpretation, following the reasoning above, would be that these states are relevant in a broader set of contexts in the wild, requiring more specific lexical associations for signals to successfully communicate about these states. In captivity, *joint travel*, *play* and *stop action* may appear in more specific contexts that constrain the interpretation of any signal being used, leading to a broader range of signals being associated with a particular state.

Yet, these interpretations remain speculative until we move beyond the signals being used and also include information about the context in the model. Nevertheless, we think the comparison of the different lexica highlights the potential of the modelling approach outlined here and foreshadows insights into how relational and contextual aspects of communicative interactions shape the corresponding means.

Discussion

Our goal in this paper was to connect the study of human and nonhuman pragmatics. We introduced a basic computational model that formalizes the process of great ape communication as signalling game based on a shared semantic lexicon. The focal LAM model is able to capture naturalistic communicative interactions between orangutans. We see this as an important starting point to deploy the modelling tool box developed by computational linguists in recent years to study great ape communication and theorize about potential evolutionary trajectories connecting it with human communication. We illustrated the productive nature of this approach by estimating and comparing lexica for different groups of orangutans (mother-offspring vs. other dyads, wild vs. captive). The results of this comparison provide at least suggestive evidence that relational and contextual factors shape the lexicon used for communication.

The present study contributes to the ongoing discussion concerning the origins of great ape gestures and their meaning. For example, Graham et al. (2018) posit that bonobo and chimpanzee gestures overlap in meaning due to a shared phylogenetic origin. Our model introduces a quantitative framework for evaluating such claims. Specifically, it enables a direct quantification and comparison of the similarity in signal meanings between multiple species by leveraging data from both signalers and receivers.

A critical extension of the model will be the inclusion of additional information sources. Bohn et al. (2022) argued that gestures are only one part of a communicative interaction and in order to understand (and predict) the outcome of the interaction one needs to consider other signals like vocalizations or facial expressions but – crucially – also the social and relational context. In linguistics and psychology, the latter is often referred to as the common ground shared between interlocutors (Clark, 1996). Previous experimental work found that great apes modify their gestural communication based on shared interactions with humans (Bohn et al., 2016). In the present study, the differences in signal meanings between mother-offspring and other dyads could be seen as further suggestive evidence that the social-interactional history between individuals directly affects how signals are being used by orangutans. For examples, mother-offspring dyads concentrated on relatively fewer messages when grooming in comparison to other dyads. One way to interpret this would be that grooming in mother-offspring pairs is less predictable because it happens in a broader range of contexts, in which it is more difficult to infer that an individual wants to (be) groom(ed), requiring more specific signals. Crucially, the framework we introduce here could be extended to directly model and quantify the importance of these information sources.

Our study has clear limitations. For example, our models were based on data from both Bornean and Sumatran orangutans, which were pooled for the present study. Since Sumatran orangutans are more gregarious and socially tolerant than their Bornean counterparts (Weingrill et al., 2011), future work should investigate whether this species difference is reflected in the inferred lexica.

Taken together, we believe that the novel computational approach introduced in this paper is a starting point for a more comprehensive approach to studying primate communication. By making concrete and transparent assumptions about potential processes behind great-ape gestural communication, implications of these assumptions can be explored computationally, thereby allowing to address empirically also recently popular philosophical ideas on gradual evolutionary trajectories from nonhuman to human communication systems (e.g. Moore, 2018; Armstrong, to appear). Furthermore, our modeling approach explicitly connects to contemporary models of human communication, allowing for formal theorizing and hypothesis testing concerning aspects of communication that may appear uniquely human.

References

Amphaeris, J., Blumstein, D. T., Shannon, G., Tenbrink, T., & Kershenbaum, A. (2023). A multifaceted framework to establish the presence of meaning in non-human communication. *Biological Reviews*, *98*(6), 1887–1909.

Armstrong, J. (to appear). The evolutionary foundations of common ground. In B. Geurts & R. Moore (Eds.), *Evolutionary pragmatics.* Oxford University Press.

Bohn, M., Call, J., & Tomasello, M. (2016). The role of past interactions in great apes' communication about absent entities. *Journal of Comparative Psychology*, *130*(4), 351.

Bohn, M., Liebal, K., Oña, L., & Tessler, M. H. (2022). Great ape communication as contextual social inference: a computational modelling perspective. *Philosophical Transactions of the Royal Society B*, *377*(1859), 20210096.

Brochhagen, T., Franke, M., & van Rooij, R. (2018). Co-evolution of lexical meaning and pragmatic use. *Cognitive Science*, *42*(8), 2757–2789.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2016, in press). Stan: A probabilistic programming language. *Journal of Statistical Software*.

Cartmill, E. A. (2023). Overcoming bias in the comparison of human language and animal communication. *Proceedings of the National Academy of Sciences*, *120*(47), e2218799120.

Cho, I.-K., & Kreps, D. M. (1987). Signaling games and stable equilibria. *The Quarterly Journal of Economics*, *102*(2), 179-221.

Clark, H. H. (1996). *Using language.* Cambridge university press.

Corballis, M. C. (2002). *From hand to mouth: The origins of language.* Princeton University Press.

Dennett, D. C. (1983). Intentional systems in cognitive ethology: The "panglossian paradigm" defended. *Behavioral and Brain Sciences*, *6*(3), 343–355.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Franke, M., & Degen, J. (2023). The softmax function: Properties, motivation, and interpretation. doi: 10.31234/osf.io/vsw47

Franke, M., & Jäger, G. (2016). Probabilistic pragmatics, or why bayes' rule is probably important for pragmatics. *Zeitschrift für sprachwissenschaft*, *35*(1), 3–44.

Fröhlich, M., Bartolotta, N., Fryns, C., Wagner, C., Momon, L., Jaffrezic, M., ... van Schaik, C. P. (2021a). Multi-component and multisensory communicative acts in orang-utans may serve different functions. *Communications Biology*, *4*(1), 917.

Fröhlich, M., Bartolotta, N., Fryns, C., Wagner, C., Momon, L., Jaffrezic, M., ... van Schaik, C. P. (2021b). Orangutans have larger gestural repertoires in captivity than in the wild—a case of weak innovation? *iScience*, *24*(11).

Fröhlich, M., Wittig, R. M., & Pika, S. (2019). The ontogeny of intentional communication in chimpanzees in the wild. *Developmental Science*, *22*(1), e12716.

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.

Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377–388.

Hawkins, R. X. D., Franke, M., Frank, M. C., Smith, K., Griffiths, T. L., & Goodman, N. D. (2023). From partners to populations: A hierarchical bayesian account of coordination and convention. *Psychological Review*, *130*(4), 977–1016.

Hobaiter, C., & Byrne, R. W. (2011). Serial gesturing by wild chimpanzees: its nature and function for communication. *Animal cognition*, *14*, 827–838.

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

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.

Lewis, D. (1969). *Convention. a philosophical study.* Cambridge, MA: Harvard University Press.

Liebal, K., Call, J., & Tomasello, M. (2004). Use of gesture sequences in chimpanzees. *American Journal of Primatology: Official Journal of the American Society of Primatologists*, *64*(4), 377–396.

Liebal, K., & Oña, L. (2018). Different approaches to meaning in primate gestural and vocal communication. *Frontiers in Psychology*, *9*, 478.

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

Moore, R. (2016). Meaning and ostension in great ape gestural communication. *Animal Cognition*, *19*(1), 223–231.

Moore, R. (2018). Gricean communication, language development, and animal minds. *Philosophy Compass*, *13*(12), e12550.

Schöller, A., & Franke, M. (2017). Semantic values as latent parameters: Testing a fixed threshold hypothesis for cardinal readings of *few* & *many*. *Linguistic Vanguard*, *3*(1).

Schuster, S., & Degen, J. (2020). I know what you're probably going to say: Listener adaptation to variable use of uncertainty expressions. *Cognition*, *203*, 104285.

Scontras, G., Tessler, M. H., & Franke, M. (2021). A practical introduction to the rational speech act modeling framework. *arXiv preprint arXiv:2105.09867*.

Scott-Phillips, T. C. (2010). Animal communication: insights from linguistic pragmatics. *Animal Behaviour*, *79*(1), e1–e4.

Skyrms, B. (2010). *Signals: Evolution, learning, and information*. Oxford: Oxford University Press.

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

Taylor, D., Clay, Z., Dahl, C. D., Zuberbühler, K., Davila-Ross, M., & Dezecache, G. (2022). Vocal functional flexibility: what it is and why it matters. *Animal Behaviour*, *186*, 93–100.

van Tiel, B., Franke, M., & Sauerland, U. (2021). Probabilistic pragmatics explains gradience and focality in natural language quantification. *Proceedings of the National Academy of Sciences*, *118*.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistical Computing*, *27*, 1413—1432.

Wärneryd, K. (1993). Cheap talk, coordination, and evolutionary stability. *Games and Economic Behavior*, *5*(4), 532–546.

Weingrill, T., Willems, E. P., Zimmermann, N., Steinmetz, H., & Heistermann, M. (2011). Species-specific patterns in fecal glucocorticoid and androgen levels in zoo-living orangutans (pongo spp.). *General and Comparative Endocrinology*, *172*(3), 446–457.

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.