Standardizing turn-taking metrics: A methodological toolkit for data annotation across species and taxa

Jolinde M.R. Vlaeyen[1,a*], Filipa Abreu[1,a], Kayla Kolff[1], Bas van Boekholt[1], Lara M. Southern[1], Pharisemène Tibesar[1], Ray Wilkinson[2] & Simone Pika[1*]

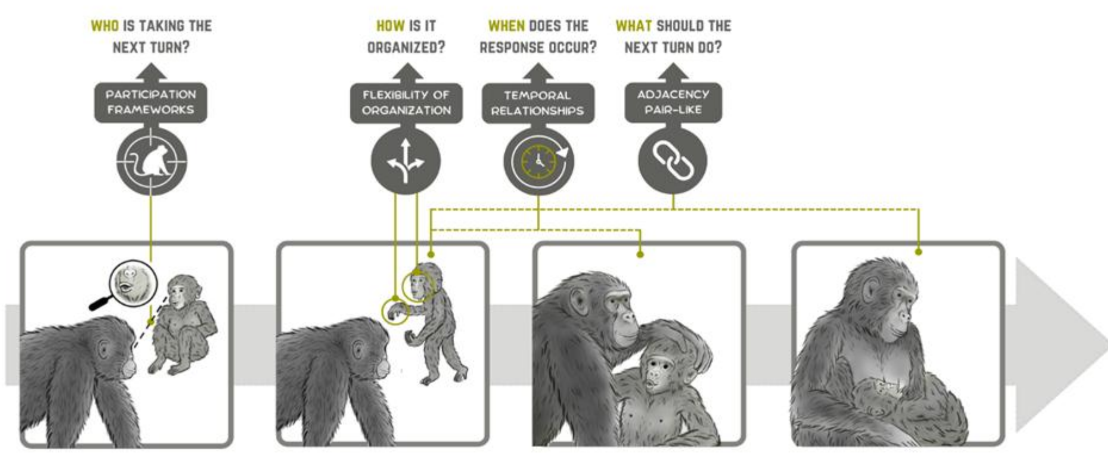
[1]Comparative BioCognition, Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

[2]Division of Human Communication Sciences, School of Allied Health Professions Nursing and Midwifery, University of Sheffield, Sheffield, United Kingdom

[a]Contributed equally

*Corresponding authors: Jolinde Vlaeyen: jvlaeyen@uos.de; Simone Pika: simone.pika@uni-osnabrueck.de

Graphical abstract



Abstract

Conversational turn-taking, the cooperative reciprocal exchange of short and flexible turns, is fundamental for communication and social coordination. Initially thought to be uniquely human, recent research showed some evidence of cooperative turn-taking also in other animal species. However, systematic evaluations and comparisons of turn-taking skills within and across species pose considerable challenges due to different measures applied and research foci on specific elements. Here, we build on a comparative framework resting on four key elements of human conversational turn-taking - flexibility, participation frameworks, temporal relationships, and adjacency pair-like sequences - to enable standardized and feasible annotation procedures across species and taxa. Proposed parameters are described in detail, and their operationalization is outlined through step-by-step annotation schemes and real-world examples. By understanding which elements of turn-taking are shared across different model systems and taxa and by mapping similarities and differences, this standardized "turn-taking toolkit" will aid in fostering a more biocentric perspective on the evolution of communication and gaining a better understanding of the evolutionary origins of language and cooperation.

Keywords: social interaction; research methods; language evolution; communication; cooperation

1. Introduction

One of the most distinctive features of human cooperation is its complexity compared to other species, enabling individuals to engage in highly structured social interactions (e.g., Apicella & Silk, 2019; Burkart et al., 2009; Melis & Semmann, 2010). These interactions are frequently accompanied by conversational turn-taking (hereafter turn-taking) - classically used to describe the alternating production of vocalizations using a variety of linguistic and non-linguistic cues between two or more individuals (Holler et al., 2015; Sacks et al., 1974). This fundamental mechanism has been suggested to be the foundation of most human interactions that regulates exchanges and facilitates coordination between interactants' mutual understanding (Levinson, 2006). Moreover, it has been suggested to be governed by universal properties that regulate turn allocation (whose turn it will be), turn transitions (when the next turn can be made and how to avoid overlap), turn size (duration of a specific turn; Clark, 1996; Sacks et al., 1974), and turn content (sequence organization; e.g., Schegloff, 2007). Beyond its structural role, turn-taking is highly cognitively demanding with human response times during conversation averaging between 0-500ms, matching the duration of an eyeblink (Stivers et al., 2009). This fast timing suggests that interactants listen to and process the acoustic information while simultaneously planning a response (Levinson & Torreira, 2015).

At the beginning of the 21st century, Levinson (2006) proposed that turn-taking is part of a package of underlying propensities in human communication, including face-to-face interactions that afford the use of gesture and gaze, and the motivation and interest in other minds, termed the *interaction engine*. Turn-taking may represent the most ancient mechanisms of the layered system of language and thus may have predated language. Support for this hypothesis stems from infant studies showing that cooperative turn-taking skills precede gestural and linguistic competence (so-called "proto conversations"), and play a key role in language development (e.g., Bruner, 1975; Clark, 2009; Ginsburg & Kilbourne, 1988), studies suggesting universality of turn-taking across cultures and languages (e.g., Stivers et al., 2009) as well as evolutionary precursors across different clades of primates (e.g., Levinson, 2016), raised the question of its biological bases and evolutionary trajectory. Recently, comparative researchers thus started to test whether turn-taking may be the evolutionary "missing link" bridging linguistic and non-linguistic species (e.g., meerkats (*Suricata suricatta*): Demartsev et al., 2018; starlings (*Sturnus vulgaris*): Henry et al., 2016; African elephants (*Loxodonta africana*): O'Connell-Rodwell et al., 2012; marmosets (*Callithrix jacchus*): Takahashi et al., 2013). (For overviews of other phenomena such as antiphonal calling, duetting and chorusing, please also see Ravignani and colleagues (2019) and Vanderhoff and Hoverud (2022)). To date, studies across insects, anurans, birds, and mammals revealed features reminiscent of human turn-taking - such as the existence of distinct time windows, adjacency pair-like sequences, overlap avoidance, and cues for turn allocation (for reviews, see: Abreu & Pika, 2022; Pika et al., 2018; Pougnault et al., 2020; van Boekholt et al., 2025).

Despite the findings, meaningful comparisons of evolutionary precursors of human turn-taking elements and shared features have been constrained by variation across employed definitions, methodologies and applied measures (Pika et al., 2018). Specifically, studies currently face three major constraints: (i) the lack of uniformity concerning the use of terms and criteria of turn-taking across and within scientific fields hampering valid cross-species comparisons (Abreu & Pika, 2022; Pika et al., 2018); (ii) the limited number of studies investigating the hallmarks of human turn-taking, which have been biased toward single elements (e.g., temporal relationships) and modalities (e.g., vocal exchanges). Since communication is inherently multimodal (e.g., Holler & Levinson, 2019; McNeill, 1992) these biases have hampered a comprehensive understanding of existing turn-taking systems, its evolutionary trajectory, and systematic, quantitative, and direct comparisons with human conversational turn-taking; (iii) the existing research bias toward nonhuman primates, limiting our understanding of whether turn-taking is a species-specific adaptation to social or environmental requirements or whether turn-taking is a phylogenetically conserved trait maintained across lineages through common descent (e.g., Ord et al., 2021). Consequently, the disparate use of definitions and measures has made direct, quantitative comparisons between human and non-human turn-taking systems exceptionally difficult, necessitating a more standardized and inclusive methodological approach.

To tackle these issues and foster systematic comparisons, Pika and colleagues (2018) pinpointed the four most crucial hallmarks of human conversational turn-taking (Sacks et al., 1974) and designed a comparative framework. It addresses:

1) *Flexibility of turn-taking organization* (how is it organized?) This element concentrates on the variety of size and ordering turns and the ability to adjust signals and actions (see Table 1 for definitions) in an interaction in case of a lack of a response.

2) *Participation framework* (who is taking the next turn?). This element focuses on the initiation of turns (who can or should produce the next signal) and techniques for attributing the next turn as the interaction is ongoing.

3) *Temporal relationships* (when do response turns occur?). This element addresses the time window between two turns during an interaction.

4) *Adjacency pair-like sequences* (what should the next turn do?). This element consists of a minimum of two turns by different participants, where a first pair part (e.g., requests) is followed by a predictable response in the form of a second pair part (e.g., grants).

Despite their aim of translating human linguistic complexity into applicable universal behavioral parameters, the field still lacks a standardized and easy-to-use method for annotation and comparing all four elements simultaneously. Most studies continue to rely on different metrics or focus on single elements, limiting robust cross-species comparisons and limiting our understanding of the evolutionary roots of coordinated interactions (but see van Boekholt et al., 2025 for an overview).

Therefore, we present a refined, operationalized toolkit for investigating the underlying elements of turn-taking across the animal kingdom and provide a revised, expanded definition of turn-taking compared to Pika and colleagues' (2018). It now refers to "the orderly exchange of communicative signals and/or behaviors between at least two individuals, governed by coordination principles that regulate turn transfer". The coordination of turn transfer should follow species-specific principles that vary in form (flexibility), sequence (contextual fit, adjacency pair-like structures), duration (temporal relationships), and distribution (turn allocation). In animals, turn-taking systems can involve various forms of exchanges, such as action-action (e.g., A grooms B, B grooms A), action-signal (e.g., A grooms B, B PRESENTS to A), signal-action (e.g., A PRESENTS to B, B grooms A), or signal-signal (e.g., A PRESENTS to B, B PRESENTS to A) interactions. While this is most often not researched in human studies (for exceptions, please see Drew & Couper-Kuhlen, 2014; Drew & Kendrick, 2018), our framework explicitly incorporates behavioral social actions alongside signals.

Unlike traditional methods that focused almost exclusively on acoustic exchanges (e.g., antiphonal calling), our approach recognizes that turn-taking can manifest as signal-signal, action-action, or mixed-modality interactions (e.g., a gesture followed by a vocalization, or reciprocal grooming; Kolff & Pika, 2025; Tibesar et al., 2026; van Boekholt & Pika, 2025). By defining turn-taking through coordination principles rather than specific modalities, this toolkit allows for the systematic annotation of all types of social exchanges, such as play, grooming, food sharing, mating, and other cooperative exchanges, under the same analytical lens as conversation.

Beyond moving from current theoretical frameworks to a universal toolkit, our aim was threefold. First, while previous studies investigated predominantly single elements (e.g., temporal relationships), here we introduce robust and feasible metrics for capturing the complete spectrum of turn-taking. Secondly, we demonstrate how these measures can be reliably applied across various communicative modalities as well as across taxa. Third, unlike theoretical frameworks or empirical research, we provide a detailed annotation guide with step-by-step processes and real-life examples to facilitate usage and implementation by other researchers.

Crucially, this toolkit is presented here as a methodological instrument for empirical testing rather than a claim of universal presence. Our goal is to provide the necessary metrics to distinguish between random behavioral successions and structured, coordinated turn-taking exchanges, acknowledging that not all animals may exhibit the same elements or degree of turn-taking. By assessing each component of turn-taking independently, the toolkit allows researchers to build a detailed profile for each species - identifying which elements are present, which are absent, and to what extent the overall pattern resembles human turn-taking. This approach facilitates systematic cross-species comparisons, enabling researchers to identify both the specific features that bring a species' communication system closer to or further from human-like turn-taking, and the ecological, cognitive, or social factors that may drive these similarities and differences. In particular, this enables researchers to identify both shared evolutionary origins (homologies) among closely related species (e.g., great apes) and functional similarities (analogies) in distantly related species (e.g., cetaceans). Thus, by establishing a standardized and comprehensive framework for turn-taking, this work provides a crucial foundation for advancing comparative research on communication systems, paving the way for deeper insights into the adaptive processes and selective pressures that shape this fundamental communicative behavior.

2. Methods: Description and implementation

2.1. Annotation Scheme

Among the plethora of in-video behavioral annotation programs (e.g, Boris, Observer), we suggest using ELAN (https://archive.mpi.nl/tla/) for several reasons. Most importantly, ELAN is the predominant tool in human language and communication research, making our cross-species coding scheme directly comparable to and compatible with existing human interaction corpora. Its hierarchical tier system is also particularly well-suited to the complexity of multimodal data, allowing researchers to map overlapping behavioral information, such as vocalizations, gaze, and social actions, onto a synchronized timeline with exceptional timestamp precision. This is critical for capturing the fine-grained temporal dynamics of turn-taking and communicative exchanges.

Moreover, we introduce a custom annotation scheme, designed to enable standardized cross-species turn-taking analysis. While the custom annotation scheme was built within ELAN and takes full advantage of its hierarchical tier system - features that are not always available or equivalently implemented in other platforms -, the underlying coding logic and term definitions are in principle transferable to alternative programs with some adaptation effort. Researchers choosing to do so should be aware that certain structural elements may need to be simplified or reorganized to fit the constraints of their preferred software, and we do not provide ready-made templates for those environments. For ELAN users, we have made the template files and annotated real-life examples in ELAN available on Figshare (https://figshare.com/s/c9fe94c30ebad0ed9c7d) to facilitate immediate implementation. A comprehensive description of the coding process and definitions of all tiers and terms is provided in Section 2.

Additionally, although a full tutorial on ELAN is beyond the scope of this study, researchers can find detailed protocols for its application in behavioral coding elsewhere (https://greatapesgestures.github.io/download/; https://tinyurl.com/yvdcdpzn; Genty & Fuchs, 2023; Grund et al., 2024 supplementary work). To investigate the four elements of turn-taking, audio and video data can be used, with synchronization of both being the preferred option for multimodality purposes.

Before coding, we recommend the development of a standardized ethogram (Altmann, 1974) to define the behavioral repertoire of the target species. While the specific content of an ethogram will naturally differ across taxa, the functional classification of the behaviors should remain consistent to ensure cross-species comparisons. Specifically, researchers must clearly distinguish between gestures (communicative signals) and social actions (mechanical behaviors). Otherwise, comparisons may be biased (e.g., one may find that species A uses more social actions than species B, but while looking more into the comparing ethograms, the social action for species A was a gesture for species B; Genty et al., 2009; Pika, 2008). Furthermore, certain behaviors may transition between these categories depending on the context and physical execution (e.g., push and pull). For example, if a chimpanzee gently pushes another chimpanzee's arm to groom this, it would be annotated as a gesture, whereas if the chimpanzee pushes the other chimpanzee so that the other falls on the ground because of the force exerted, this would be annotated as an action. To address these complexities, please see Table 1 for our definition of communicative modalities and the fulfillment of primary criteria and Table S1 for how to annotate them.

Table 1. Definitions of different terms used in a turn-taking interaction.

Terms	Definition
Signals	Production of information via a stimulus (e.g., facial expressions and gestures, and vocalizations; Marler, 1967).
Facial Expressions	Expressive movements of different facial muscles, such as mouth, cheeks, lips, or eyes, to produce a visual signal (Liebal et al., 2004).
Gestures	Movements of the extremities or head and body postures that are directed to a social partner, mechanically ineffective – meaning produced to request or reach a goal, requiring active participation of the interactant - and potentially elicit a response by the recipient (Grund et al., 2024; Pika, 2008).
Vocalizations	Sounds produced via the vibration of vocal folds of the larynx voluntarily or involuntarily (Nieder & Mooney, 2020).
Oro-facial sounds	Sounds produced with the lips and/or teeth (mouth can be opened or not), without passing air via the vocal folds.
Multi-modal signals	Fixed multi-modal signals, such as the "scream face" in chimpanzees that is always coupled with a "scream vocalization" due to production mechanics, involve obligatorily linked modalities, and free multi-modal signals, like the "pout face" that may or may not be accompanied by a vocalization, are flexibly combined and not biomechanically constrained (Partan & Marler, 2005).
Social Actions	Any socially directed embodied behavior that leads to the perceived goal through direct manipulation of another's body/body parts or the movement of one's own body (Fröhlich et al., 2016).
Interactions	Alternation of produced social actions and verbal and/or non-verbal signals between participants that co-inhabit a shared environment (Enfield & Levinson, 2006).
Turns	Signals or actions made by one individual to initiate, continue or end an interaction.

2.2. Inter-Rater reliability

To ensure consistent application of the coding scheme, we recommend conducting inter-rater reliability assessments. Individuals blind to the hypotheses should independently annotate a randomly selected subset (e.g., 15%) of the data, and agreement should be assessed using a statistic appropriate for the type of data (e.g., Cohen's Kappa for categorical data, or EasyDIAG in ELAN itself; Holle & Rein, 2015). An acceptable threshold for agreement should be established a priori (e.g., Kappa > 0.70), and discrepancies should be resolved through discussion and consensus, and coding guidelines should be refined as needed.

2.3. Turn-taking elements: Metrics applied and how to annotate them

Note: All coding examples shown are hypothetical demonstrations. Specific rules (e.g., response timing) should be established based on species-specific data.

A. Flexibility of turn-taking organization

Question: How is the interaction organized? This element refers to the variety of size and ordering of turns, markers of goal-directed behavior, and the ability to adjust, combine and/or change signals/actions in an interaction to achieve the desired goal (Abreu & Pika, 2022; Pika et al., 2018; Fig. 1&2).

Operationalized parameters: Flexibility can be assessed by quantifying response waiting, repetition (persistence and elaboration), and turn transitions utilizing the following tiers (Table S1): Vocalizations, Gestures, Facial Expressions, Actions, and Repetition. Repetition is annotated in real-time as observed, while response waiting and turn transitions are identified during the analysis phase.

(a) Response waiting - period during which an initiator pauses after producing a signal/action, anticipating a response from the recipient. The waiting period should last at least 1 second (Hobaiter et al., 2017) depending on the species. If no response is received within this time, the interaction can be annotated as "No Response".

(b) Repetition (e.g., Fröhlich & van Schaik, 2022; Roberts et al., 2013) - situations where an individual reiterates a signal/action after a pause (see definition in participation framework element) until a response is received (species-specific timing). Repetition can be categorized as:

a. Persistence: The exact same signal/action is repeated after a pause (see Fig. 1);
b. Elaboration: Another signal/action is repeated after a pause (see Fig. 2)[1].

(c) Turn transitions - identified by consecutive signal/action pairs produced by different individuals within an interaction. These transitions should be visible upon completion of the coding process (see adjacency pair-like sequences for further information).

Coding Considerations: Ambiguous responses should be discussed in detail with other annotators to determine whether it constitutes a valid response. Additionally, species-specific maximum waiting times should be determined, with particular consideration of whether the individual is monitoring (gaze and body orientation) the partner.

With these three distinct metrics, we can quantify the frequency of actions, gestures, vocalizations, facial expressions, and multifaceted combinations (any multimodal and multicomponent combinations of the above-mentioned tiers from the same individual; see Fröhlich & van Schaik, 2018; Hex & Rubenstein, 2024 for details on multimodality and multicomponent communication) during a turn-taking event. Further analysis can also explore the variation of sequences exchanged between initiator and recipient, and their respective participation roles within and over multiple interactions - for example, identifying recurring patterns such as A-B-A or A-B-C (but see Fröhlich et al., 2025 for a thorough review on flexibility). Notably, response waiting and persistence/elaboration may indicate attempts at repair - "fixing" a misunderstanding or communicative breakdown (Schegloff, 2007) - in situations where an adequate response is not immediately received (e.g., Heesen et al., 2022).

[1] There may be exceptions to these outlined cases, but see temporal relationships for further information.

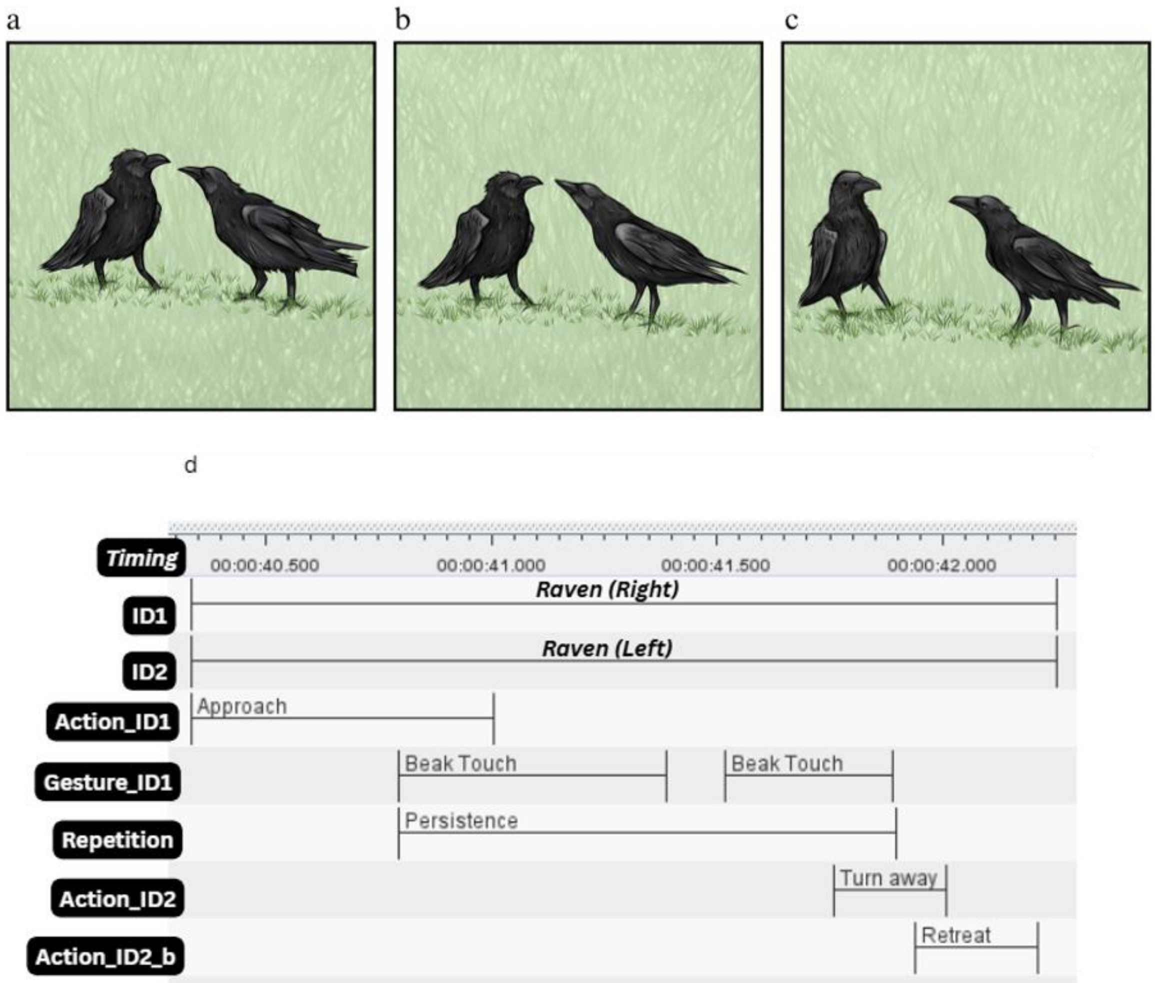


Figure 1. Flexibility element in an interaction between two ravens (*Corvus corax*): (a) ID1 approaches and tries to TOUCH ID2 with his beak; (b) when ID2 does not respond (~0.2s), ID1 repeats the gesture (persistence) while also coming closer; (c) ID2 turns her body away and retreats; (d) ELAN screenshot of the annotated interaction using the tiers from Table S1. © Luara Martins.

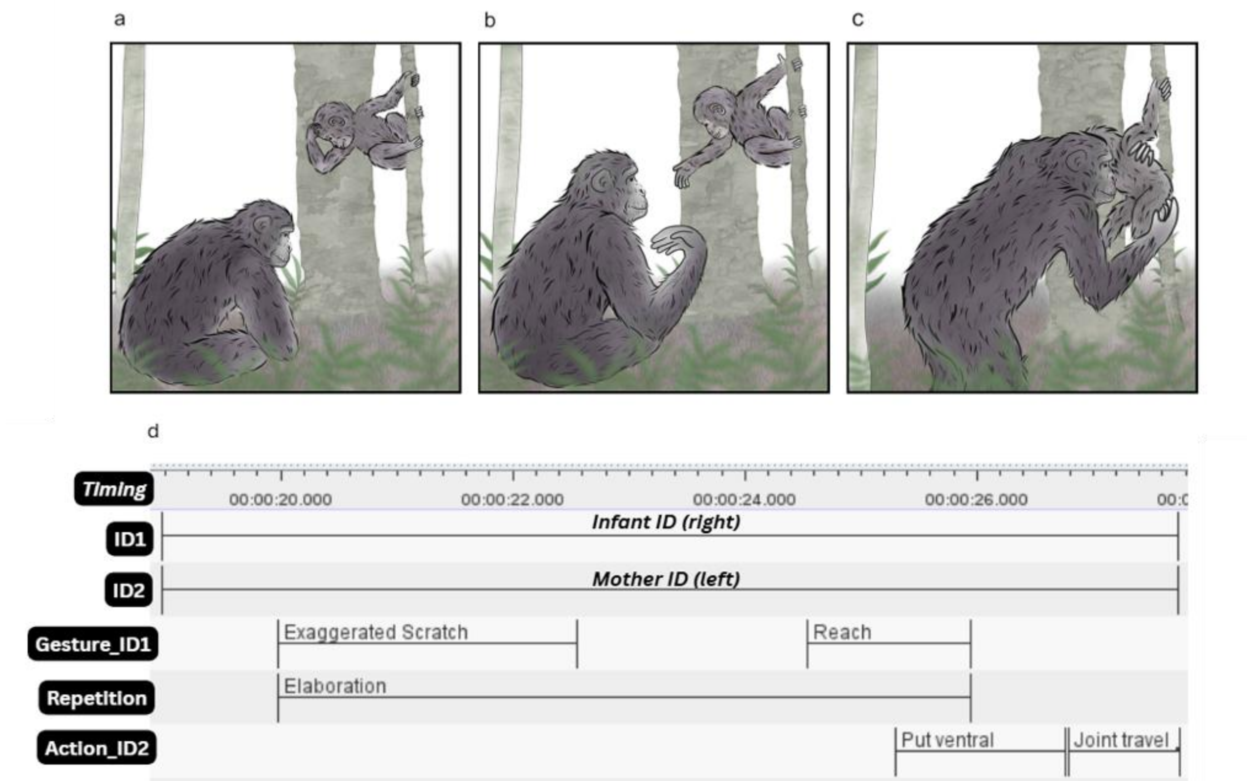


Figure 2. Flexibility element in the joint travel context in a chimpanzee (*Pan troglodytes*) mother (ID2) and her infant (ID1): (a) ID1 performs an EXAGGERATED LOUD SCRATCH toward her mother and waits; (b) after ~2s, ID1produces a REACH (elaboration); (c) ID2 responds by positioning her infant ventral for travel; (d) ELAN screenshot of the annotated interaction using the tiers from Table S1. © Luara Martins.

B. *Participation framework*

Question: Who is taking the next turn? This element focuses on the initiation of turns (who can or should produce the next signal), and the capacity and/or techniques used for attributing the next turn to another individual/audience (Pika et al., 2018; Fig. 3&4).

Operationalized parameters: Participation framework can be assessed by quantifying orientation patterns, proximity dynamics, pauses, and phase responses utilizing the following tiers (Table S1): Gaze, BodyOrientation, and Proximity. Orientation and Proximity are annotated in real-time as observed, while pauses and phase responses are identified after coding.

(a) Orientation (Fig. 3) - directional alignment of interacting individuals, encompassing both attentional state and bodily direction. Orientation can be categorized as either Gaze- or Body-directed:
   a. Gaze: Face direction is defined as the angle between the initiator´s face and the body position of the other individual. Whenever possible, focusing on the eye gaze of the initiator is preferable, as this gives finer details.
   b. Body: Shoulder and chest direction (species dependent) can be used as metrics, and defined as the angle of these areas of the initiator in relation to the other individual.

(b) Proximity - spatial distance maintained between interacting individuals throughout the interaction sequence. We recommend coding it by using categories based on the anatomy of the respective model species and the type of data collected:
   a. Video data: Using species-specific body measurements (e.g., arm's length, body length) rather than absolute metric distances (e.g., meters) may provide more reliable proximity criteria across different taxa.
   b. Audio data: Distinguish between visible and not visible proximity at the beginning of an interaction during data collection.

(c) Pauses - time elapsed (offset-onset) between two consecutive signals/actions produced by the same individual when there is no response from another individual.

(d) Phase responses — time elapsed (offset-onset) between two separate signals/actions produced by the same individual with another individual producing a signal/action in between. We recommend comparing pauses, phase responses, and response latencies (see temporal relationships) to allow the assessment of how an individual's signaling timing is influenced or adjusted by others. This comparison helps identify the mechanisms by which turns are allocated and coordinated during an interaction.

Coding Considerations: We recommend adapting the degrees of orientation to the study species and to annotate orientation within a cone shape, so that individuals are considered facing even if their gaze is directed slightly upward or downward (Fig. 3). Orientation coding applies solely to video footage and is continuously annotated from the onset (or earlier for a more detailed depiction) of an interaction until one (or both) of the interactors end the interaction, or until individuals are no longer visible. Due to the inherent constraints of animal movement, angular transitions cannot skip intermediate ranges (e.g., moving from <90° to >180° must include coding for 90°-180°). For audio-only data, pauses and phase responses can serve as the primary metric for assessing participation framework when orientation cannot be determined.

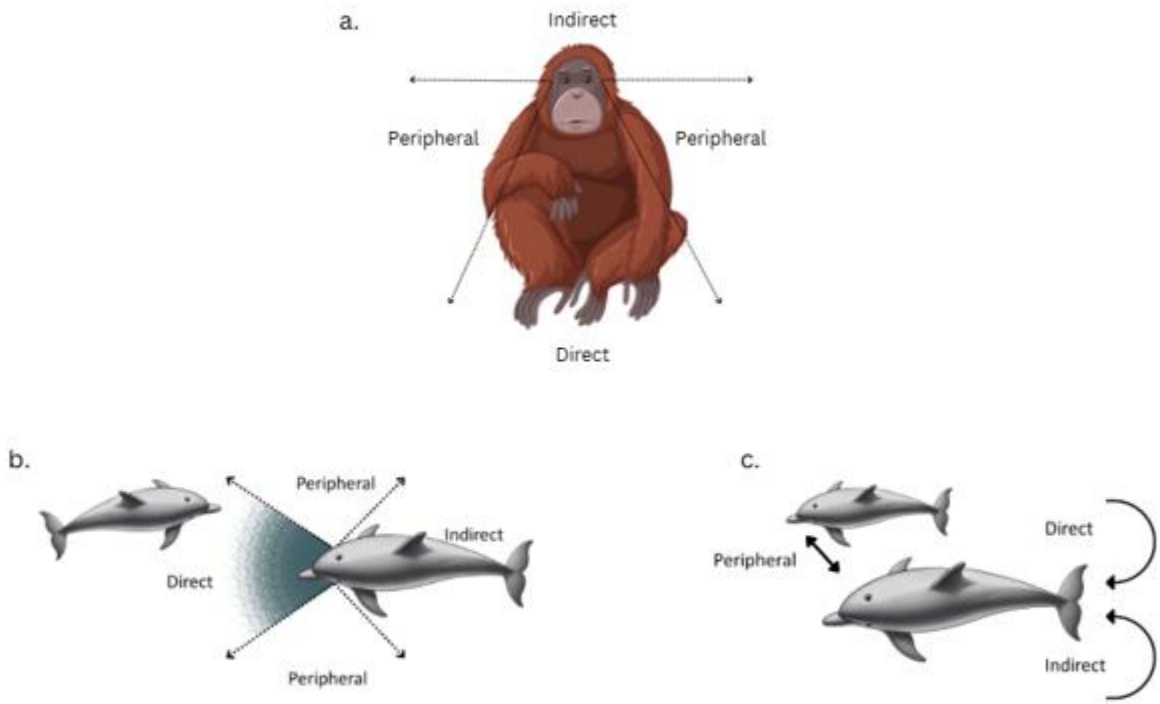


Figure 3. Illustration representing gaze direction and body orientation categories: (a) animals with front-facing eyes: <90° (Direct), 90°-180° (Peripheral), >180° (Indirect) (e.g., Pika et al., 2003); b. gaze direction animals with lateral-facing eyes: <60° (Direct), 60-200° (Peripheral), >200° (Indirect) (e.g., Connor & Smolker, 1996); and (c) body orientation for lateral-facing eyes relative to conspecific. In the illustrated position, the orientation is peripheral; if the individual rotates so that its belly faces the other, the orientation is direct, whereas turning the back toward the other indicates indirect orientation.

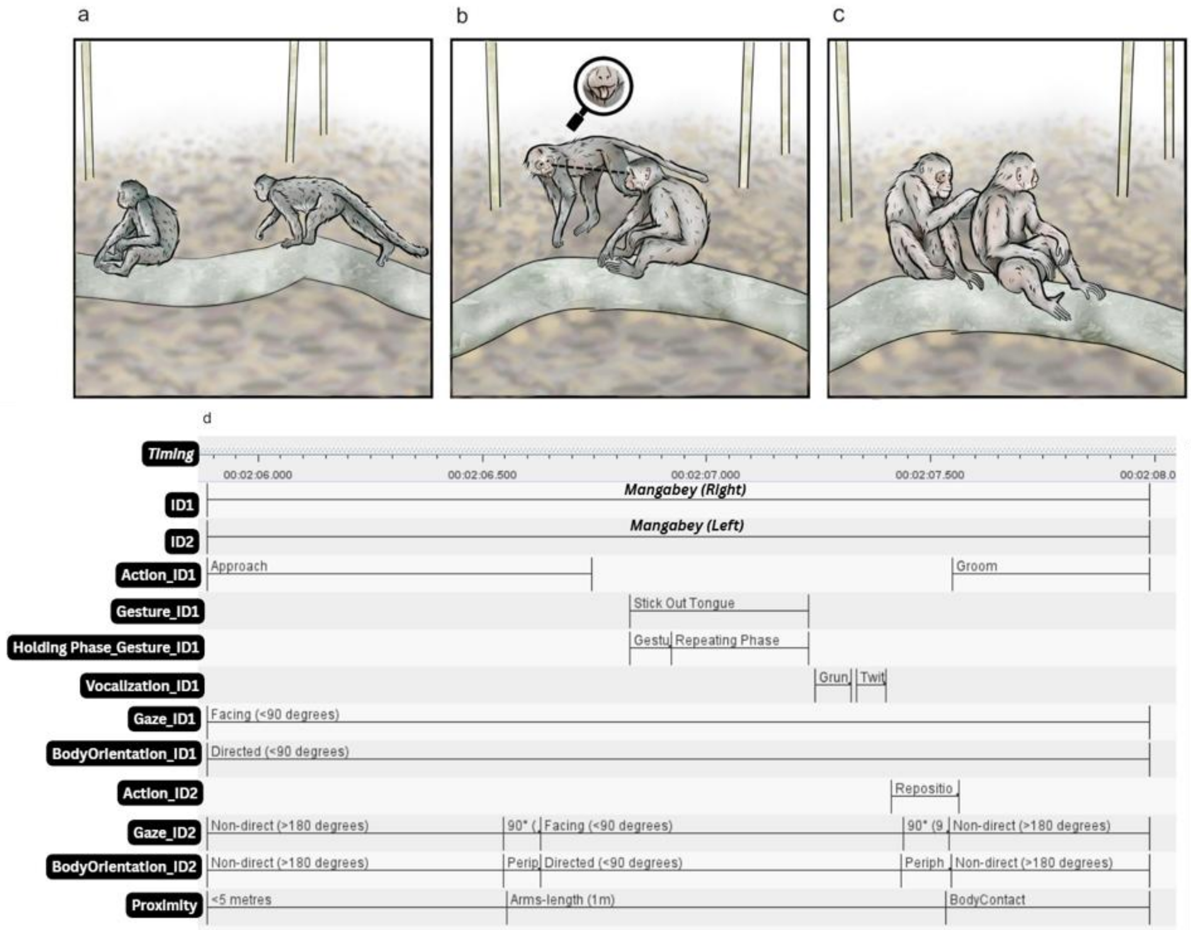


Figure 4. Participation framework element in a grooming initiation in sooty mangabeys (*Cercocebus atys*): (a) ID1 approaches within <5m (Proximity), gazes at ID2 (<90°) who does not look at ID1 (gaze direction: >180°); (b) in arm's length to ID2, ID1 directs her gaze (<90°), performs TONGUE OUT, and produces a GRUNT and a TWITTER; ID2 matches gaze and body orientation (<90°); (c) ID1 begins grooming after 0.2s; (d) ELAN screenshot of the annotated interaction using the tiers from Table S1. © Luara Martins.

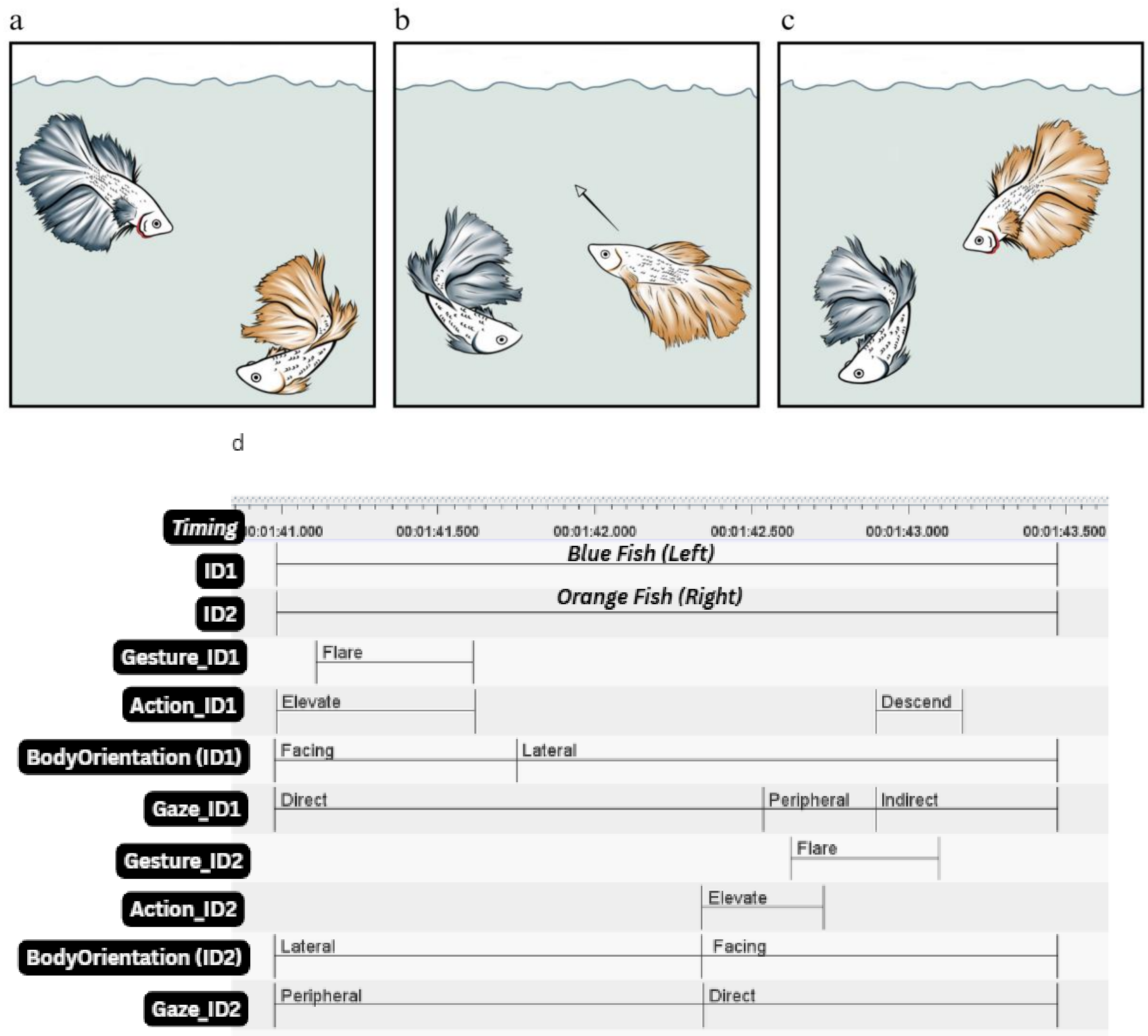


Figure 5. Participation framework element in agonistic visual displays of Siamese fighting fish (*Betta splendens*) (based on the study of Everett et al., 2025): (a) ID1 (blue fish) elevates in the water column and produces FLARE while facing forward; (b) as ID1 turns laterally, ID2 (orange fish) elevates and orients toward ID1; (c) ID2 produces FLARE while elevated and facing forward as ID1 turns away and descends in the water tank; (d) ELAN screenshot of the annotated interaction using the tiers from Table S1. © Luara Martins.

C. *Temporal relationships*

Question: When do response turns occur? This element addresses the time window between the signal/action, and its response, which can be species (Pika et al., 2018) and possibly also modality and context-specific (Fig.7).

Operationalized parameters: The temporal relationship has been the most investigated element (e.g., Pougnault et al., 2020; Ravignani et al., 2019; van Boekholt et al., 2025), and can be analyzed by quantifying response latency patterns, gap and overlap distributions, and timing variations by calculating timing differences between relevant tiers (Table S1). All temporal measurements are derived during the analysis phase by measuring time differences between annotated signals/actions across different individuals. Rather than selecting a metric arbitrarily, the choice should be guided by the research question under investigation.

(a) Response Latency - time elapsed between two consecutive signals/actions, or signal-action combinations produced by the different individuals involved, encompassing gaps and overlaps in turn transitions (this is, between turns; Fig. 6). Response latency can be categorized based on measurement approach:

   a. Onset-to-Onset: Measures the time an individual takes to "understand" a signal or action - recommended when the focus is on how promptly an individual responds in relation to the start of articulation (e.g., in humans, this would involve the cognitive processes that occur before a signal/action is fully articulated), but does not include overlap avoidance. This is analogous to measures used in human sign language research (e.g., De Vos et al., 2016), making it particularly suitable for comparisons involving species with prominent gestural communication (e.g., primates).
   b. Offset-to-Onset: Measures the time between the end of one signal/action and the start of the next - recommended when the focus is to examine the capacity of the animal to avoid overlap and respond promptly or for interactions that are mainly vocal. This is analogous to measures used in human vocal language research (e.g., Stivers et al., 2009), making it particularly suitable for comparisons involving species with prominent vocal communication (e.g., birds).
   c. Holding Phase: Measures the time from the moment the signal/action has been fully executed until the position is fully retracted (e.g., the RAISE ARM gesture in bonobos) or has received a response (e.g., the PRESENT gesture in chimpanzees for grooming purposes).

Coding Considerations: To ensure robust cross-species comparisons, researchers must adopt a consistent temporal anchor for measuring turn transitions. We have provided three standardized metrics (onset-to-onset, offset-to-onset, or holding-to-onset) to accommodate the diverse physical properties of different communicative modalities. Thus, the choice should align with the respective research objectives and the specific cognitive or behavioral processes under investigation, as well as the modalities examined. While all of the aforementioned metrics are valid, it is crucial to specify and report which metric was employed in the methods section of the resulting manuscript to allow for proper comparisons. Species-specific timing constraints should be established based on modal characteristics and ecological context. Finally, to investigate sequences of signals/actions, we suggest using the offset of the last signal/action of the sequence to the onset of the response because this method allows for an analysis of the continuity and flow of interactions, helping to understand how individuals manage the timing of their responses within a sequence.

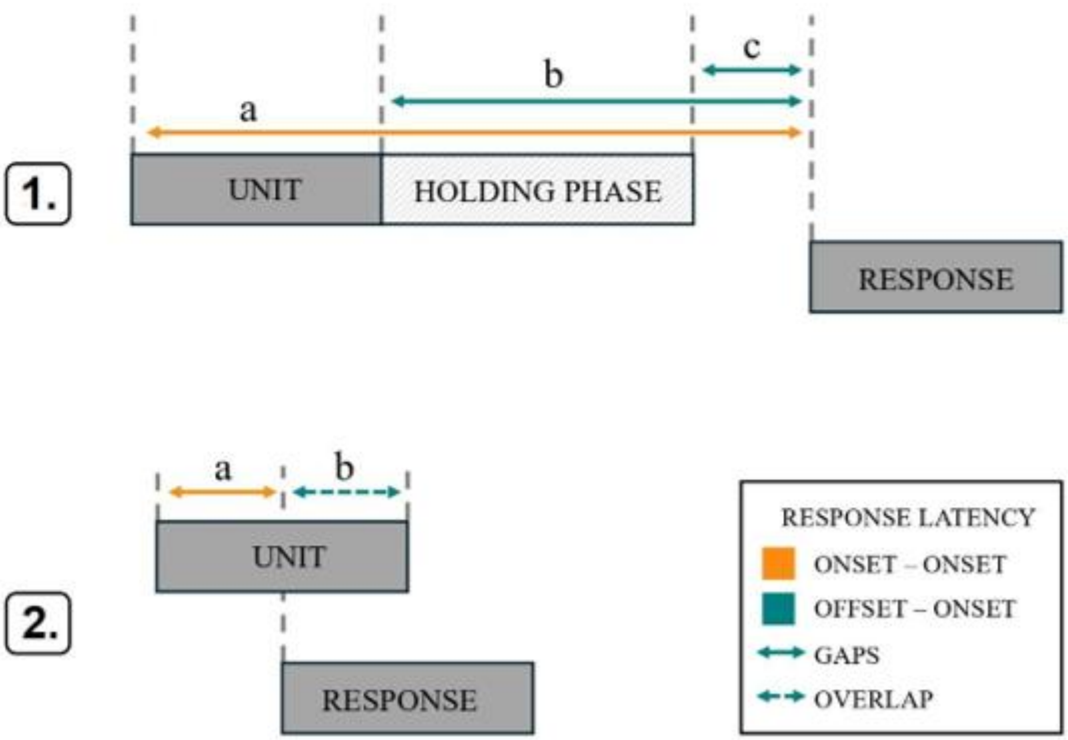


Figure 6. Diagram representing different operational ways of annotating a response latency within an interaction for a 1. non-overlapping response or 2. overlapping response: (a) unit (signal/action) onset to

response onset (the individual may respond before the signal/action is complete); (b) unit (signal/action) offset (for gestures, it is the end of fully executed gesture) to response onset encompassing gaps and overlaps; and (c) gesture holding phase offset (retraction point) to response onset. Note that this can be used for multiple modalities.

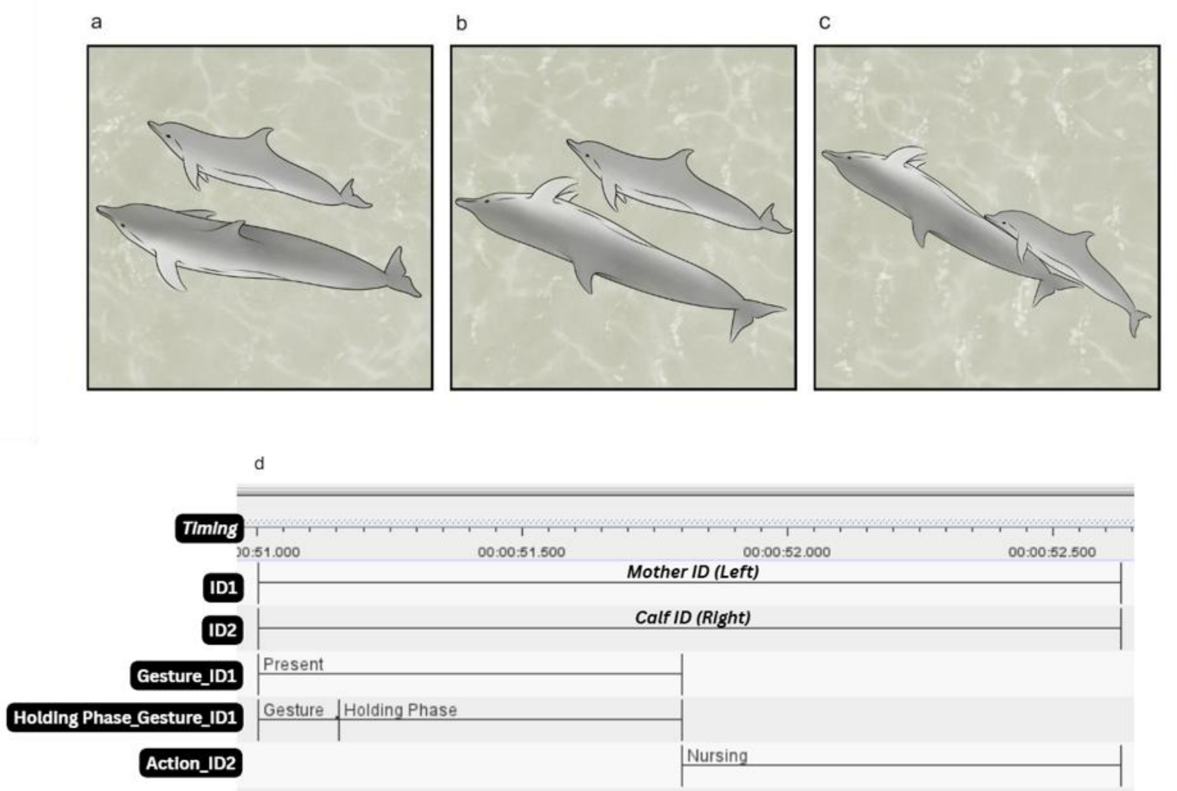


Figure 7. Temporal Relationships element in the nursing context of dolphins (*Tursiops spp*.): (a) mother (ID1) and calf (ID2) swim side by side in close proximity; (b) ID1 PRESENTS her belly to ID2; (c) ID2 begins nursing with a latency of 0.8s (onset-onset), 0s (offset-onset including holding phase), or 0.65s (offset-onset excluding holding phase); (d) ELAN screenshot of the annotated interaction using the tiers from Table S1. © Luara Martins.

D. *Adjacency pair-like sequences*

Question: What should the next turn do? This element addresses adjacency pairs (Schegloff, 2007), a core structural unit of conversation that can be recursively reproduced and expanded. It consists of a minimum of two turns by different individuals, where a *first pair part* (e.g., a request) is followed by a predictable response in the form of a *second pair part* (e.g., granting; Pika et al., 2018; Fig. 6).

Operationalized parameters: Adjacency pair-like sequences are assessed by quantifying signal-response contingencies, sequential pattern frequencies, and response appropriateness utilizing the following tiers (Table S1): Vocalizations, Gestures, Facial Expressions, Actions. Individual signals/actions are annotated during the annotation phase, while signal-response pairings and sequential patterns are identified post-hoc during the analysis phase, through sequential and collocation analyses that determine contingent non-random associations between:

(a) all the adjacent signals/actions produced by participants during the interaction.
(b) the sequences employed by the initiator and recipient.

Coding Considerations: Response appropriateness criteria must be established based on species-specific behavioral repertoires and contextual norms. Sequential analyses should employ statistical methods (e.g., lag-sequential analysis, Markov chain modeling) to distinguish predictable patterns from random associations. Minimum sequence length and timing parameters should be defined based on species-typical interaction dynamics.

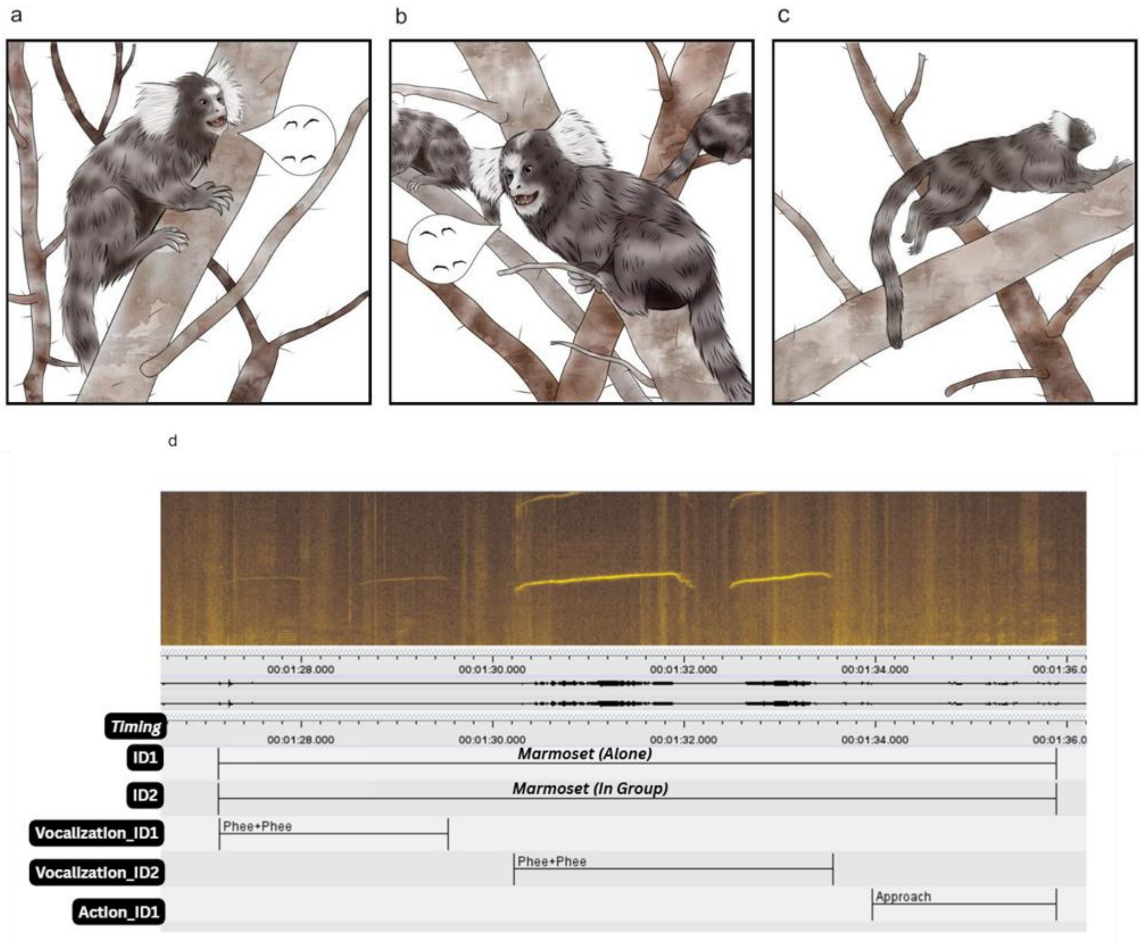


Figure 8. Adjacency pair-like sequences element in a vocal interaction of common marmosets (*Callithrix jacchus*): (a) ID1 is alone (no visual contact with his conspecifics) and emits a PHEE-PHEE; (b) ID2 from another group replies with a matching vocalization within 10s; (c) ID1 travels toward the neighboring group upon hearing the vocalization; (d) ELAN screenshot of the annotated interaction using the tiers from Table S1. © Luara Martins.

2.4. Integrating Elements: A Bonobo Mother-Infant Communication Example

Here, we aim to illustrate the utilization of our practical toolkit through a detailed example of a bonobo (*Pan paniscus*) mother-infant interaction, demonstrating how all four turn-taking elements integrate into a single communicative exchange, and how they can be systematically annotated using the ELAN annotation software (see Fig. 9 for visual representation and corresponding annotation screenshot).

Consider the following scenario involving an infant requesting to nurse. First, the infant (right) gazes at his mother (left; *Participation framework* – annotated as “<90” in the Gaze and BodyOrientation tiers), and produces a POUT FACE (FacialExpression_ID1; SignalType tier is Unimodal). The mother does not respond (although she looks at the infant; Gaze/BodyOrientation tiers), and the infant elaborates by adding an ARM RAISE (Gesture_ID1 and BodyPart_ID1) and POUT MOAN (Vocalization_ID1) to the POUT FACE (FacialExpression_ID1). These signal combinations and elaboration are part of the *Flexibility* element (Repetition: “Elaboration”; SignalType now: Multimodal). After 0.5s (*Temporal relationships* - offset-onset; see Timing), the mother responds with an OPEN ARM gesture (Gesture_ID2), and allows the infant to access her nipple. This creates an *Adjacency pair-like sequence*: The infant’s multimodal request (POUT FACE + POUT MOAN + ARM RAISE; turn one) receives an “aligning” response from the mother (OPEN ARM; turn two). This triggers further turn-transitions: The infant approaches (Action_ID1), the mother TOUCHES the infant (Gesture_ID2), and the infant nurses (Action_ID1). Additional contextual annotation captures the interaction's full complexity: The Context tier tracks the behavioral context (e.g., resting → nursing (only context visible in Fig. 9.) → playing), the Proximity tier records distance changes (e.g., arms-length → body contact), the Missing_Parts tier indicates data completeness, the External_Factors tier notes influences (e.g., community vocalizations), and the Interaction_ID enables statistical analysis across exchanges. In sum, this systematic annotation demonstrates how our toolkit captures turn-taking complexity within natural communicative exchanges, demonstrating both the depth of information available and its practical implementation

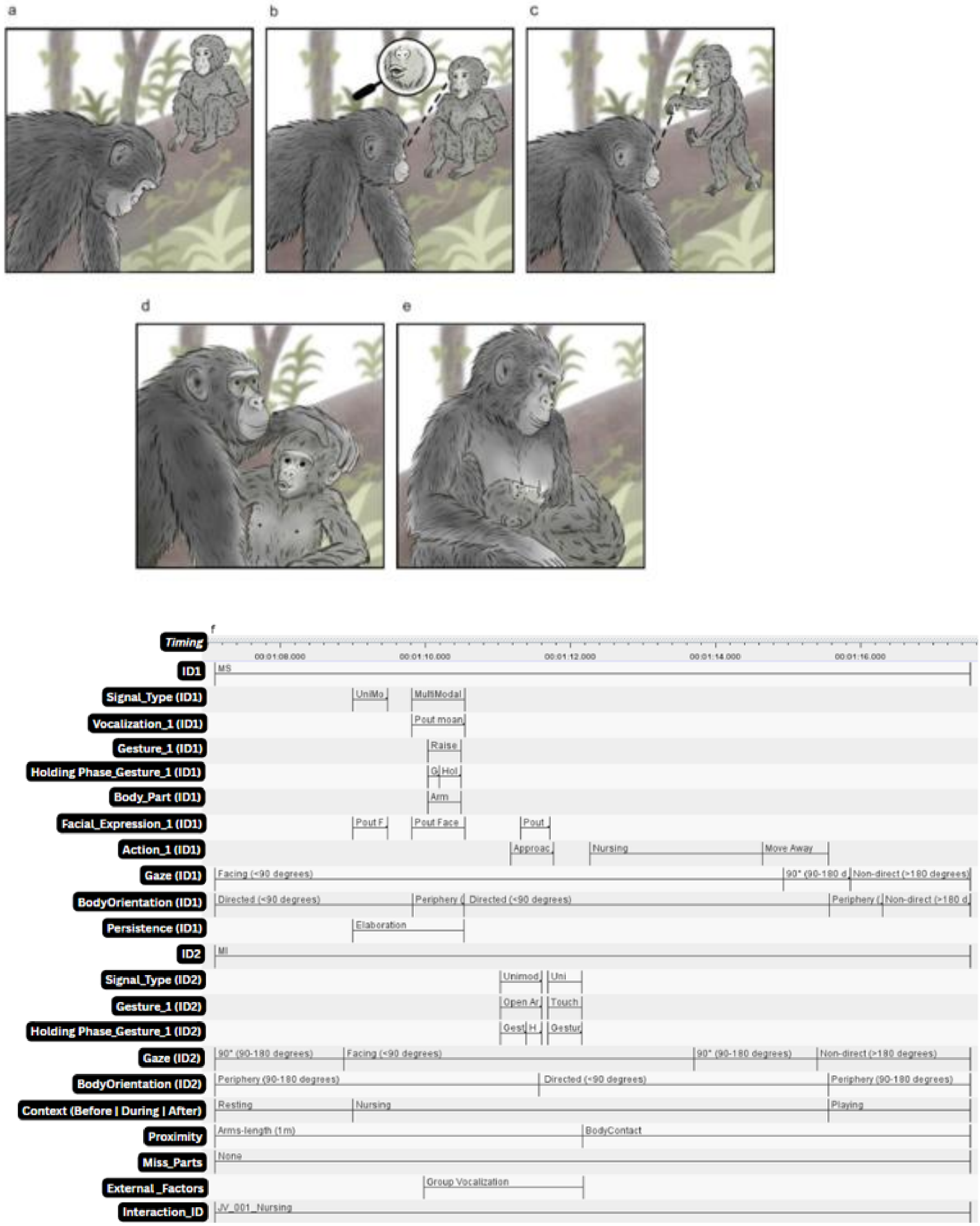


Figure 9. All turn-taking elements in a nursing interaction between a bonobo infant (ID1) and its mother (ID2): (a) ID1 gazes at ID2; (b) when ID2 returns gaze, ID1 produces a POUT FACE; (c) ID2 does not respond, so ID1 also employs an ARM RAISE and a POUT MOAN in addition to the POUT FACE; (d) ID1

responds with an OPEN ARM; (e) the request for nursing has been granted; (f) ELAN screenshot of the annotated interaction using the tiers from Table S1. © Luara Martins.

3. Future Directions

Our operationalized turn-taking toolkit advances the study of communicative evolution by shifting the focus from isolated, top-down linguistic categories to a multidimensional, bottom-up analysis of interaction. By integrating and operationalizing the four core elements proposed by Pika and colleagues (2018), we move beyond single elements and modality biases that have historically been applied in this field and have hampered informed comparisons. This bottom-up approach allows to deconstruct complex social exchanges into their fundamental building blocks, such as flexibility and adjacency pair-like sequences, and to identify the behavioral components of cooperation and turn-taking that may have preceded formal language. Crucially, our toolkit highlights that focusing on any single element does not capture the intricate system and involved complexity of turn-taking. For instance, although the temporal rule of overlap avoidance (Sacks et al., 1974), offers valuable insight into timing exchanges, it captures only one dimension of turn-taking and falls short of characterizing the full complexity of human-like conversation (Dingemanse & Enfield, 2024).

By broadening the analytical scope to include social actions alongside signals, our toolkit captures the functional logic of interaction that precedes and supports formal language. This comprehensive approach is an essential step for identifying meaningful similarities and differences between animal exchanges and human social action during conversation (e.g., Fröhlich et al., 2016; Habib-Dassetto et al., 2023; Kolff & Pika, 2025; Levinson, 2019; Mondémé, 2022; van Boekholt & Pika, 2025). Furthermore, this multidimensional perspective provides the empirical rigor needed to resolve a longstanding debate: whether fast-paced, extended social exchanges are uniquely human (e.g., Melis et al., 2016; Tomasello, 2010) represent an ancient mechanism characterizing cooperative social interactions found across diverse taxa (Levinson & Holler, 2014; Levinson, 2016). Where previous cross-species research has been constrained by methodological inconsistencies and a narrow focus on single modalities (Pika et al., 2018), our systematic toolkit offers the standardized, modality-agnostic metrics necessary for definitive comparisons across the animal kingdom.

Although non-human primates remain a favored model system for investigating evolutionary precursors of language and factors triggering its evolution (e.g., Levinson, 2016), our approach stands out by offering the possibility of applying our methodology across a wide range of taxa and communication modalities to allow us to pinpoint exactly which species exhibit organization most akin to human turn-taking, regardless of the modality used. Such comparisons enable us to assess the role of ecological and social factors on communicative abilities and complexity (e.g., group size, cooperation demands, predation pressure) that also may shape turn-taking. Furthermore, as Sacks and colleagues (1974) already noted, humans use multiple turn-taking systems, including those for coordinated actions. Studying these different turn-taking systems in non-linguistic species is the only way to empirically disentangle communication-specific elements of human turn-taking from the general coordination mechanisms shared across social animals, and thus to identify what, if anything, is truly unique to human conversation.

Overall, this toolkit enables us to construct a hierarchy of resemblance to human turn-taking, distinguishing between similarities driven by shared ancestry and those arising through convergent evolution (e.g., Henry et al., 2016). For instance, if distantly related species such as crows exhibit turn-taking structures similar to chimpanzees, it may indicate that these behaviors converge functionally rather than reflect a shared evolutionary origin, a distinction with profound implications for understanding the selective pressures that shape communicative complexity. Finally, including interspecies interactions can also further deepen our insight into the fundamental principles of sociality and emergent meaning that transcend strictly linguistic or species-specific norms (Mondémé, 2022). By revealing how turn-taking systems evolve across the animal kingdom, this work brings us closer to answering one of the most fundamental questions in cognitive science: what makes human language - and the social intelligence that underlies it - truly unique.

4. Conclusion

In summary, broad comparative analyses will reveal how turn-taking systems evolve and whether they lay the groundwork for more sophisticated communication across the animal kingdom. Our toolkit enables systematic and comparable data collection within and across species and taxa while maintaining analytical flexibility to determine turn-taking presence in their specific contexts. It is important to note that, although all of the presented metrics are valid, researchers should report and specify which metrics were employed for transparency and easier cross-comparisons in the future. Lastly, broadening the scope of turn-taking research will advance the understanding of communication evolution and its interplay with cognition, sociality, and cooperative behavior, moving us closer to unraveling the roots of language.

Acknowledgements

This research was funded by an EU-Consolidator grant (772000, TurnTaking) to Simone Pika of the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program. We are indebted to Antonio Souto, Catherine Crockford, David Watts, Geraldo Baracuhy, John Mitani, Kevin Langergraber, Martin Surbeck, Nicola Schiel, Roman Wittig and Stefanie Gazda for allowing access to their field sites and animal populations to collect the observational data involving the examples presented in this manuscript.

Author contributions

Conceptualization: Simone Pika, Jolinde Vlaeyen, Filipa Abreu; **Funding acquisition:** Simone Pika; **Methodology:** Simone Pika, Jolinde Vlaeyen, Filipa Abreu, Kayla Kolff, Bas van Boekholt, Lara Southern, and Pharisemène Tibesar; **Supervision:** Simone Pika; **Writing – original draft:** Jolinde Vlaeyen and Filipa Abreu; **Writing – review and editing:** all authors.

Data Availability Statement

Data is available at https://figshare.com/s/c9fe94c30ebad0ed9c7d.

Conflict of Interest statement

We have no conflicts of interest to declare.

4. References

Abreu, F., & Pika, S. (2022). Turn-taking skills in mammals: A systematic review into development and acquisition. *Frontiers in Ecology and Evolution*, *10*, 987253. https://doi.org/10.3389/fevo.2022.987253

Altmann, J. (1974). Observational study of behavior: sampling methods. *Behaviour, 49*(3), 227-266. https://doi.org/10.1163/156853974X00534

Apicella, C. L., & Silk, J. B. (2019). The evolution of human cooperation. *Current Biology*, *29*(11), R447-R450. https://doi.org/10.1016/j.cub.2019.03.036

Bruner, J. S. (1975). The ontogenesis of speech acts. *Journal of child language*, *2*(1), 1-19. https://doi.org/10.1017/S0305000900000866

Burkart, J. M., Hrdy, S. B., & Van Schaik, C. P. (2009). Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, *18*(5), 175-186. https://doi.org/10.1002/evan.20222

Clark, E. V. (2009). *First Language Acquisition*. Cambridge University Press. https://books.google.nl/books?id=fVV9aC4FeWoC

Clark, H. H. (1996). *Using language*. Cambridge university press. https://doi.org/10.1017/CBO9780511620539

Connor, R. C., & Smolker, R. A. (1996). 'Pop'goes the dolphin: A vocalization male bottlenose dolphins produce during consortships. *Behaviour*, *133*(9-10), 643-662.

De Vos, C., Torreira, F., & Levinson, S. C. (2016). *Turn-timing in signed conversations: coordinating stroke-to-stroke turn boundaries* (Vol. 4).

Demartsev, V., Strandburg-Peshkin, A., Ruffner, M., & Manser, M. (2018). Vocal turn-taking in meerkat group calling sessions. *Current Biology*, *28*(22), 3661-3666. e3663.

Dingemanse, M., & Enfield, N. (2024). Interactive repair and the foundations of language. *Trends in cognitive sciences*. https://doi.org/10.1016/j.tics.2023.09.003

Drew, P., & Couper-Kuhlen, E. (2014). *Requesting in social interaction* (Vol. 26). John Benjamins Publishing Company.

Drew, P., & Kendrick, K. H. (2018). Searching for trouble: Recruiting assistance through embodied action. *Social Interaction. Video-Based Studies of Human Sociality*.

Enfield, N. J., & Levinson, S. C. (2006). Roots of human sociality. *New York: Berg*.

Everett, C. P., Norovich, A. L., Burke, J. E., Whiteway, M. R., Villamayor, P. R., Shih, P.-Y., Zhu, Y., Paninski, L., & Bendesky, A. (2025). Coordination and persistence of aggressive visual communication in Siamese fighting fish. *Cell reports*, *44*(1).

Fröhlich, M., Kuchenbuch, P., Müller, G., Fruth, B., Furuichi, T., Wittig, R. M., & Pika, S. (2016). Unpeeling the layers of language: Bonobos and chimpanzees engage in cooperative turn-taking sequences. *Scientific reports*, *6*(1), 1-14.

Fröhlich, M., & van Schaik, C. P. (2018). The function of primate multimodal communication. *Animal cognition*, *21*(5), 619-629.

Fröhlich, M., & van Schaik, C. P. (2022). Social tolerance and interactional opportunities as drivers of gestural redoings in orang-utans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1859).

Genty, E., Breuer, T., Hobaiter, C., & Byrne, R. W. (2009). Gestural communication of the gorilla (*Gorilla gorilla*): repertoire, intentionality and possible origins. *Animal cognition*, *12*, 527-546. https://doi.org/10.1007/s10071-009-0213-4

Genty, E., & Fuchs, M. (2023). GApS: A Coding Scheme for Great Apes Signals in ELAN (v1.0.2). https://doi.org/10.5281/zenodo.7573664

Ginsburg, G. P., & Kilbourne, B. K. (1988). Emergence of vocal alternation in mother-infant interchanges. *Journal of child language*, *15*(2), 221-235.

Grund, C., Badihi, G., Graham, K. E., Safryghin, A., & Hobaiter, C. (2024). GesturalOrigins: A bottom-up framework for establishing systematic gesture data across ape species. *Behavior research methods*, *56*(2), 986-1001. https://doi.org/10.3758/s13428-023-02082-9

Habib-Dassetto, L., Lemasson, A., Portes, C., & Montant, M. (2023). Discontinuist and continuist approaches of language evolution… and beyond. *Revue de primatologie*(14).

Heesen, R., Fröhlich, M., Sievers, C., Woensdregt, M., & Dingemanse, M. (2022). Coordinating social action: a primer for the cross-species investigation of communicative repair. *Philosophical Transactions of the Royal Society B*, *377*(1859), 20210110. https://doi.org/10.1098/rstb.2021.0110

Henry, L., Craig, A., Lemasson, A., & Hausberger, M. (2016). Social coordination in animal vocal interactions. Is there any evidence of turn-taking? The starling as an animal model. *H., Casillas, M., Levinson, SC, eds.(2016). Turn-Taking in Human Communicative Interaction. Lausanne: Frontiers Media. doi: 10.3389*, *4*, 39.

Hex, S. B., & Rubenstein, D. I. (2024). Using networks to visualize, analyse and interpret multimodal communication. *Animal behaviour*, *207*, 295-317.

Hobaiter, C., Byrne, R. W., & Zuberbühler, K. (2017). Wild chimpanzees' use of single and combined vocal and gestural signals. *Behavioral ecology and sociobiology*, *71*(6), 96.

Holle, H., & Rein, R. (2015). EasyDIAg: A tool for easy determination of interrater agreement. *Behavior research methods*, *47*, 837-847.

Holler, J., Kendrick, K. H., Casillas, M., & Levinson, S. C. (2015). Turn-taking in human communicative interaction. In (Vol. 6, pp. 1919): Frontiers Media SA.

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in cognitive sciences*, *23*(8), 639-652.

Kolff, K., & Pika, S. (2025). Turn-taking in grooming interactions of chimpanzees (Pan troglodytes schweinfurthii) in the wild: the role of demographic and social factors. *Animal cognition*, *28*(1), 26.

Levinson, S. (2019). Interactional foundations of language: The interaction engine hypothesis. In *Human language: From genes and brain to behavior* (pp. 189-200). MIT Press.

Levinson, S., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130302. https://doi.org/10.1098/rstb.2013.0302

Levinson, S., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in psychology*, *6*, 731. https://doi.org/10.3389/fpsyg.2015.00731

Levinson, S. C. (2006). On the human "interaction engine". In N. J. Enfield & S. C. Levinson (Eds.), *Roots of Human Sociality: Culture, Cognition and Interaction* (pp. 39-69). Berg.

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends in cognitive sciences*, *20*(1), 6-14. https://doi.org/10.1016/j.tics.2015.10.010

Liebal, K., Pika, S., & Tomasello, M. (2004). Social communication in siamangs (Symphalangus syndactylus): use of gestures and facial expressions. *Primates*, *45*, 41-57.

Marler, P. (1967). Animal Communication Signals: We are beginning to understand how the structure of animal signals relates to the function they serve. *Science*, *157*(3790), 769-774.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* University of Chicago Press. https://doi.org/

Melis, A. P., Grocke, P., Kalbitz, J., & Tomasello, M. (2016). One for you, one for me: Humans' unique turn-taking skills. *Psychological science*, *27*(7), 987-996.

Melis, A. P., & Semmann, D. (2010). How is human cooperation different? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1553), 2663-2674. https://doi.org/10.1098/rstb.2010.0157

Mondémé, C. (2022). Why study turn-taking sequences in interspecies interactions? *Journal for the Theory of Social Behaviour*, *52*(1), 67-85. https://doi.org/10.1111/jtsb.12295

Nieder, A., & Mooney, R. (2020). The neurobiology of innate, volitional and learned vocalizations in mammals and birds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789).

O'Connell-Rodwell, C., Wood, J., Wyman, M., Redfield, S., Puria, S., & Hart, L. (2012). Antiphonal vocal bouts associated with departures in free-ranging African elephant family groups (Loxodonta africana). *Bioacoustics*, *21*(3), 215-224.

Ord, T. J., Klomp, D. A., Summers, T. C., Diesmos, A., Ahmad, N., & Das, I. (2021). Deep-time convergent evolution in animal communication presented by shared adaptations for coping with noise in lizards and other animals. *Ecology letters*, *24*(9), 1750-1761. https://doi.org/10.1111/ele.13773

Partan, S. R., & Marler, P. (2005). Issues in the classification of multimodal communication signals. *The American Naturalist*, *166*(2), 231-245.

Pika, S. (2008). What is the nature of the gestural communication of great apes. *The shared mind*, 165-186. https://doi.org/10.1075/celcr.12.11pik

Pika, S., Liebal, K., & Tomasello, M. (2003). Gestural communication in young gorillas (*Gorilla gorilla*): gestural repertoire, learning, and use. *American Journal of Primatology: Official Journal of the American Society of Primatologists*, *60*(3), 95-111. https://doi.org/10.1002/ajp.10097

Pika, S., Wilkinson, R., Kendrick, K. H., & Vernes, S. C. (2018). Taking turns: bridging the gap between human and animal communication. *Proceedings of the Royal Society B*, *285*(1880), 20180598. https://doi.org/10.1098/rspb.2018.0598

Pougnault, L., Levréro, F., Mulot, B., & Lemasson, A. (2020). Breaking conversational rules matters to captive gorillas: A playback experiment. *Scientific reports*, *10*(1), 1-10. https://doi.org/10.1038/s41598-020-63923-7

Ravignani, A., Verga, L., & Greenfield, M. D. (2019). Interactive rhythms across species: the evolutionary biology of animal chorusing and turn-taking. *Annals of the New York Academy of Sciences*, *1453*(1), 12-21. https://doi.org/10.1111/nyas.14230

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

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction* (pp. 7-55). Elsevier. https://doi.org/10.1353/lan.1974.0010

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis I* (Vol. 1). Cambridge university press. https://doi.org/10.1017/CBO9780511791208

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., De Ruiter, J. P., & Yoon, K.-E. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587-10592. https://doi.org/10.1073/pnas.0903616106

Takahashi, D. Y., Narayanan, D. Z., & Ghazanfar, A. A. (2013). Coupled oscillator dynamics of vocal turn-taking in monkeys. *Current Biology*, *23*(21), 2162-2168.

Tibesar, P., Crockford, C., Le Floch, A., & Pika, S. (2026). Turn-taking in grooming interactions of wild sooty mangabeys (Cercocebus atys). *Animal cognition*.

Tomasello, M. (2010). *Origins of human communication*. MIT press. https://doi.org/10.7551/mitpress/7551.001.0001

van Boekholt, B., Kolff, K., Vlaeyen, J. M., Abreu, F., Southern, L. M., Tibesar, P., & Pika, S. (2025). Turn-Taking Beyond Humans: Insights From Nonhuman Animals.

van Boekholt, B., & Pika, S. (2025). Infrastructure of mother-infant interactions across development in chimpanzees (Pan troglodytes) in the wild. *Evolution and Human Behavior*, *46*(2), 106671.

Vanderhoff, E. N., & Hoverud, N. B. (2022). Perspectives on antiphonal calling, duetting and counter-singing in non-primate mammals: an overview with notes on the coordinated vocalizations of bamboo rats (*Dactylomys spp.*, Rodentia: Echimyidae). *Frontiers in Ecology and Evolution*, *10*, 906546. https://doi.org/10.3389/fevo.2022.906546