Sequence organization and timing of bonobo mother-infant interactions

Federico Rossano
Max Planck Institute for Evolutionary Anthropology, Germany

In recent years, some scholars have claimed that humans are unique in their capacity and motivation to engage in cooperative communication and extensive, fast-paced social interactions. While research on gestural communication in great apes has offered important findings concerning the gestural repertoires of different species, very little is known about the sequential organization of primates’ communicative behavior during interactions. Drawing on a conversation analytic framework, this paper addresses this gap by investigating the sequential organization of bonobo mother-infant interactions, and more specifically, how individuals solicit carries from one another. It shows how bonobos establish participation frameworks before producing a carry request gesture and how the ensuing communicative actions can be organized in adjacency-pair sequences. Moreover, the timing between the initiation of an action and its response is similar to what has been documented in adult human interaction. Finally, it outlines some of the orderly practices bonobos use to deal with the absence of response from the addressed participants in carry sequences.

Keywords: adjacency pair; pan paniscus; conversation analysis; gestures; interactional time; sequence organization

1. Introduction

In recent years, some scholars have claimed that the capacity and motivation to engage in cooperative communication and to engage systematically in fast-paced and extended social interactions are uniquely human skills (e.g. Levinson 2006; Tomasello 2008). The question addressed in this paper is to what extent this might be the case. I will address this question by outlining the structure and timing of bonobo (Pan paniscus) mother-infant interactions in a specific sequential environment, pick up and carry sequences, and by comparing them with the structure and timing of sequences of human social interaction.
Over the last thirty years, much has been written on gestural and vocal communication in non-human primates, partly aimed at documenting the extent and flexibility of their repertoire, and partly aimed at comparing these repertoires with the evolution of language in humans. Gestures in great apes have captured the bulk of scholarly attention since, for the most part, they appear to be learned, used flexibly – contrary to most vocalizations (Call & Tomasello 2007; Pollick & de Waal 2007, but see, e.g. Crockford et al. 2012 for evidence of some flexibility in these as well) – and are sensitive to audience. Great apes monitor their recipients’ attention before producing a visual gesture, and, if their recipient is distracted, they will either use attention getters to elicit attention (see, e.g. Hostetter, Hopkins & Cantero 2001; Tomasello, Gust & Frost 1989) or they will move in the recipient’s line of sight to facilitate perception of the visual signal (Liebal, Pika, Call & Tomasello 2004). In addition, when great apes produce a gesture, they tend to wait for a response from their recipient (e.g. Plooj 1978; Tomasello, George, Kruger, Farrar & Evans 1985). While previous research has described this communicative behavior as displaying signs of both goal-directedness (Bruner 1971) and communicative intention (Hobaiter & Byrne 2011b; Sperber & Wilson 1986; but see Tomasello 2008 for an opposing perspective), I suggest that it also displays that some form of response is expected and, therefore, that the relevance of a response from a specific individual has been established.

2. **Gestures in great apes: Form and function**

Previous studies on gestural communication in great apes have focused on establishing the gestural “lexicon” of each species (i.e. the number and type of different gestures) based on specific individuals or groups (e.g. Call & Tomasello 2007). Traditionally, analysts have identified behavioral forms and inferred their functions by looking at their general context of use. However, because what constitutes a gesture is more broadly defined in this line of research than in studies of human gestures (i.e. the term includes not only visible behavior but also tactile and auditory signals), assessing what counts as a gesture, as opposed to a non-communicative physical action, can be problematic. In addressing this issue, Liebal and Call (2012) have suggested that visible gestures and physical actions are not discrete categories but may in fact be arranged on a continuum. Specifically, they have proposed certain features to help determine precisely where certain behaviors may lie on this continuum between action and gestures (e.g. mechanical ineffectiveness, gazing at each other, waiting for a response).

Just as categorizing body movements as gestures is not straightforward, neither is characterizing contexts of use, and thus, the functions of gestures.
In previous studies, contexts of use are usually categorized as “sex, travel, nursing, feeding, play, agonistic [e.g. fighting], affiliative [e.g. grooming]” (e.g. Call & Tomasello 2007; Genty et al. 2009; Pika, Liebal & Tomasello 2005), and the claim is that all social actions of great apes can be characterized according to these categories/functions. Some researchers have gone even further to suggest that the only social action great apes perform is requesting (Cartmill & Byrne 2010). This simplification arises by starting from forms and inferring functions, rather than proceeding in the opposite direction, that is, starting from the specific activities primates engage in and the social goals they try to achieve, and then identifying how those activities get initiated and negotiated via embodied communicative behavior. Recent work has begun to move in this direction by focusing on primates’ behavior within a singular activity: co-locomotion via ventral or dorsal carry initiated and executed by bonobo mother-infant dyads (Halina, Rossano & Tomasello 2013; Hutchins & Johnson 2009). These studies report the behaviors of both mother and infant in achieving successful carries. With successful practices to elicit carries identified, it is possible to examine cases in which individuals engage in those practices but fail to achieve the same outcome. Furthermore, by starting from activities (as opposed to forms), we are not only able to identify the range of practices that elicit those activities, but we may also begin to understand why certain practices are implemented over others in specific situations. From this perspective, the question of whether or not a given practice constitutes a gesture or a physical action becomes less central, and instead, the formation and recognition of a given practice, as enacted via the embodied behavior of primates, comes to the fore. Ultimately, adopting this view allows us to investigate gesture sequences at a different level of granularity.

3. Gesture sequences in great apes

Previous research on gesture sequences has aimed at identifying something similar to sentence structure in the communication of chimpanzees (Liebal, Call & Tomasello 2004; Hobaiter & Byrne 2011a) and gorillas (Genty & Byrne 2010). These studies did not find any combination of gestures comparable to a human sentence, nor did they find that great apes systematically produce attention getters before initiating the core gesture of a sequence. They also did not find that using multiple gestures in a sequence increases the likelihood of obtaining a response. However, Liebal, Call and Tomasello (2004) found that one third of all gestures belonged to a gesture sequence, two-thirds of gesture sequences were comprised of only two gestures, just less than half of all gesture sequences were repeats of the same gesture, and, most gesture sequences occurred when the first element of the
sequence did not generate a response. Neither Liebal, Call and Tomasello (2004) nor Hobaiter and Byrne (2011a) found any systematic ordering of gestures in a sequence in terms of their relative efficacy (e.g. the gesture most likely to obtain a response first, and the least effective ones afterward). Hobaiter and Byrne (2011a) also reported that older individuals more frequently produced single gestures and sequences of gestures tended to occur when the first one failed. Given this, they argued that producing a sequence of gestures is a ‘fail-safe’ strategy for younger individuals to increase the chance of successful outcomes. Indeed, producing multiple gestures ultimately allowed for the production of the ‘effective’ types at some point in the sequence. As a result, these young individuals learn which gestures are more effective. In general, these studies suggest that effectiveness in eliciting a response is a general property of the gesture, one that is not tied to other contextual configurations. The expectation is that gesture selection is tied to its relative efficacy, although none of these studies report any clear evidence for this perspective.

One possible explanation for the amount of negative findings reported in these studies may have to do with their definition of gesture sequences. According to Liebal et al. (2004: 382), “two gestures were considered to be part of a sequence if one sender performed them toward the same recipient and in the same context [i.e. feeding, playing, etc.] within five seconds of each other.” Genty and Byrne (2010) and Hobaiter and Byrne (2011a) employed the same definition but with a timing of only one second. However, without conducting a systematic investigation of the timing of these sequences in great apes, these time scales appear arbitrary, and may in fact be too long or too short in duration. Also, by adopting these definitions of gesture sequences, these studies collapsed gestures that are in fact related by two different types of organization: sequential organization and sequence organization (Schegloff 2007: 2). Sequential organization refers to the relative positioning of items with respect to one another, as in a temporal or spatial organization where one unit/object occurs after another. Sequence organization, on the other hand, is a special type of sequential organization – it considers the intrinsic relationship between units/objects that occur sequentially (e.g. units that belong to a single course of action or activity). For example, two social actions that occur one after the other in the same general situation (e.g. eating) may accomplish different courses of action and have no relationship to one another (e.g. asking someone for something to drink, and, one second later, inviting that same person to a party). These two social actions have a sequential relation but they are not part of the same sequence. In such cases, there is no expectation that actions are ordered because they are not part of the same course of action; they are not pursuing the same goal. On the other hand, if someone asks for something to drink and the other person does not give her/him something, then asking again (even in a slightly different way), is part of the same course of action because that individual
is pursuing an absent response. It is crucial to recognize these two organizations as
distinct in order to avoid treating actions that are not related as though they are,
and correspondingly, to avoid assuming that they are necessarily deployed in an
orderly way. Drawing on these two conversation analytic organizations, in what
follows I investigate gesture sequences in mother-infant bonobo interactions in a
way that treats them as distinct. First, however, a brief overview of the conversation
analytic framework for investigating these issues is required.

4. The sequential organization of action

In conversation analysis (CA), producing and understanding communicative
moves during human social interaction have been labeled the action formation
problem (from the speaker’s point of view), and the action recognition/ascription
problem (from the recipient’s point of view), respectively. These problems concern
the resources participants deploy in social interaction to make their actions (e.g.
requests, offers, etc.) recognizable and intelligible to co-participants (Schegloff
2007; but for a critical review, see Levinson 2013). Conversation analysts have long
been interested in how social actions can be (made) intelligible, as intelligibility
is required to achieve mutual understanding, which provides for the successful
engagement of cooperative interactions. A major contribution to that endeavor
has been outlining the sequential organization of social action via turns at talk.
The adjacency pair is the most fundamental structure in this organization, one
that can be recursively reproduced and expanded in conversation (e.g. Sacks 1992;
Schegloff 2007). In its minimal, unexpanded form, an adjacency pair has the fol-
lowing features:

a. It is composed of two turns
b. by different participants
c. that are adjacently placed
d. and are relatively ordered into first pair parts (FPPs) and second pair parts
   (SPPs). FPPs are actions in first position that initiate some exchange
   (e.g. questions, requests, offers), and SPPs are actions in second position that
   are responsive to FPPs (e.g. answers, grants, acceptances).
e. These components are pair-type related; that is, not every SPP can follow any
   FPP, so that, for example, a greeting makes relevant a reciprocal greeting in
   response, a question – an answer, etc. (Schegloff 2007:13).

The same utterance in first or second position can implement very different kinds
of actions. Saying “it's snowing” in first position can count as a noticing or an
announcement, while in response to “do you want to go for a walk?” (i.e. in second
position) would count as a rejection. Accordingly, the sequential position of any action in an adjacency pair plays an integral role in both action recognition and ascription – a role as important as the design of the action.

The timing between a sequence initiating action (FPP) and a responsive one (SPP) is also meaningful from a conversation analytic perspective. Preferred responses are ones that are produced without delay (e.g. Heritage 1984), which in adult human conversations usually means within 0.2 seconds of the completion of an FPP (Stivers et al. 2009). Dispreferred responses are usually marked by delay, hesitations, explanations, etc. The absence of response, or delays longer than 0.2 seconds, can indicate a recipient’s trouble in understanding a speaker’s FPP (Pomerantz 1984), and one possible remedy for such trouble is for the speaker to pursue a response by producing another FPP (e.g. Pomerantz 1984). In ordinary conversation, speakers of an FPP appear to wait the same amount of time for a recipient to respond to an initial FPP or subsequent pursuits, and this is usually less than 1 second (Rossano 2012a).

In situations that involve multiple participants, establishing who is part of the interaction and who is not (and who just happens to be a bystander), who is being addressed and who should respond is important. In conversations among humans, parties frequently establish and negotiate participation frameworks (e.g. Goffman 1981; Goodwin & Goodwin 2004), which achieve precisely these functions. One way in which a participation framework can be achieved spatially is by establishing an F-formation. According to Kendon (1977) an F-formation is an arrangement of participants’ bodies that provides equal and exclusive access to the space between them. It can be used to communicate the availability of participants (e.g. for whom that communication is designed), and also to maintain a common focus of attention. When one party looks at another, s/he not only monitors the other participant’s attention but also ensures that the recipient knows that s/he is being addressed and that a response is expected from her/him. While CA has fruitfully applied these concepts to human social interaction, in this paper I provide evidence of the establishment of participation frameworks, the existence of adjacency pair structures and pursuit practices in bonobo mother-infant interactions and the similarity between the timing of their actions and the timing of human social actions in conversation.

5. Data

The following examples come from a larger project developed by Halina, Rossano and Tomasello (2013) that analyzed over 400 hours of video recordings of ten different mother-infant bonobo dyads from six different zoos. Specifically, Halina
et al. investigated these data in terms of actions and gestures leading to mother-infant carries and showed that, in each dyad, infants had developed a variable number of gesture types through ontogenetic ritualization (i.e. gestures are not biologically predetermined but rather progressively ritualized during ontogeny via repeated dyadic interaction, e.g. Tomasello et al. 1997) to get mother to engage in a carry. The fact that these gestures were ontogenetically ritualized means that the selection of gestures deployed by infants to elicit a carry by the mother was specifically recipient designed for their mothers, with whom they had ritualized them (Garfinkel 1967; Sacks, Schegloff & Jefferson 1974). Note, that even though the gestures were recipient designed, it does not mean that other individuals might not be able to infer or learn their meaning, especially if they can observe the gestures and the reactions they obtain repeatedly because of living in the same environment.

Engaging in a carry in non-human primates can be considered a joint action (Clark 1996), whereby each participant plays a specific role in bringing the action about. For example, the infant usually holds onto mother’s body while mother moves, which is different from human infants. Nonetheless, a carry does not always play out as a ratified joint action. Indeed, mothers can force infants into a carry by grabbing them and holding them onto their venter with one arm. The opposite situation is not possible; that is, an infant cannot force the mother into a carry. This makes it particularly interesting for investigating asymmetric roles in these interactions because if the infant wants to move, then s/he is posed with the problem of how to get the mother to engage in a carry. In general, infants in their second year of life do not move further than a few meters away from their mothers; thus, carries are the means through which infants can move greater distances. Indeed, they are a frequent means for co-locomotion through their local environment (on average, they occur five times per hour, Halina et al. 2013).

In this paper, I focus on carry sequences performed by two mother-infant dyads (Yasa-Fimi and Ulindi-Loto). Each dyad was videotaped using focal animal sampling with the infant serving as the focal subject (Martin & Bateson 2007). The two infants, Fimi (female) and Loto (male) are half-siblings, as they have the same father (Kuno). Both infants were born and live in Leipzig zoo (Germany), were raised by their own mothers and were recorded during their second year of life (Table 1). Yasa is Fimi’s mother, while Ulindi is Loto’s mother.

Table 1. Subject information

<table>
<thead>
<tr>
<th>Infant</th>
<th>Mother</th>
<th>Infant sex</th>
<th>Infant age (years, months)</th>
<th>Mother age (years, months)</th>
<th>Hours recorded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fimi</td>
<td>Yasa</td>
<td>Female</td>
<td>1.3–1.9</td>
<td>12.2–12.8</td>
<td>102</td>
</tr>
<tr>
<td>Loto</td>
<td>Ulindi</td>
<td>Male</td>
<td>1.2–2</td>
<td>17.1–17.11</td>
<td>133</td>
</tr>
</tbody>
</table>
In the following analysis, the sequence-initiating move (i.e. the first sequential attempt to elicit a carry) will be considered the base FPP. On the other hand, the behavioral response that makes the carry possible will be considered the SPP (for the mother this usually involves approaching the infant, placing a hand/arm behind the infant’s back, slightly scooping her up, and getting up from the ground before the carry is actually performed).

In what follows, I first present a fine-grained analysis of the timing relationship between a FPP and a SPP in a carry sequence by presenting examples of adjacency pair sequences in which a sequence-initiating action by either the infant of the mother (a FPP; e.g. a ‘request’) is immediately followed by a sequentially appropriate response (a SPP; i.e. engagement in a carry). Then I consider how the individual who attempts to elicit a carry handles the absence of an appropriate SPP or some other unexpected responsive behavior by producing pursuits. Finally, I offer some quantitative findings concerning some of the qualitative claims made in this discussion.

6. Adjacency pairs and participation framework

In the first 4 examples, I show how mothers’ “invitation” to engage in a carry and infants’ “requests” to be carried can be produced via a single gesture and how the addressed recipients immediately respond to these FPPs by engaging in a carry. In the first Example (1) Fimi requests Yasa to pick her up and carry her via raising her left arm and bending her wrist.

(1) Wrist bent. Fimi.

![Figure 1a–c. Fimi wrist bent to Yasa](image)

In Figure 1a, Fimi looks towards her mother Yasa, who is eating leaves nearby. Fimi waits until her mother looks towards her and then produces a wrist bent gesture by lifting her left arm and bending her wrist. I suggest that this is a FPP as this gesture constitutes Fimi’s request to get carried by Yasa and Yasa treats the gesture...
as such. Indeed, the mother immediately abandons eating, gets up, places her right hand behind Fimi’s back and then they engage in a ventral carry (the SPP). Note some features related to the articulation of this gesture: it is a visible gesture and as such it has to be produced at some distance from the addressee. At the same time, it is important not only that the addressee can perceive Fimi’s gesture, i.e. that mother is attending to the gesture, as it is usually emphasized in research on primate communication. For the FPP (the request via a wrist bent gesture) to be successful, it is also important that the addressee (Yasa) knows that this gesture is aimed at herself and therefore that a response is expected from her and not any other individual. Note that, although rarely, other individuals carry around Fimi at times and at the time of this recording, five other bonobos were present in the same enclosure with Fimi and Yasa. The sustained gaze by the individual producing the request (Fimi) towards the addressee expected to produce the responsive behavior (Yasa) has an ostensive dimension as it publicly and intentionally selects Yasa as the recipient of the FPP. In this first example, then, we can see the following structure:

1. Infant establishes participation framework by sustaining gaze towards mother and checking mother’s gaze (Figures 1a–b);
2. Infant produces FPP: wrist bent gesture (Figure 1b); and,
3. Mother produces SPP: immediately recognizes infant’s request and satisfies it (Figure 1c).

The timing passing between the stroke of the wrist bent gesture in 1b and the beginning of mother getting up is remarkably short: less than 0.2 seconds. As soon as the infant produces the wrist bent gesture, mother gets up and picks Fimi up to begin a carry.

The next Example (2) shows even further the importance of the gaze direction and body orientation of the producer of the FPP in establishing the relevance of a response by a specific individual.

(2) Spread legs. Loto.

In this example Loto makes his way along the rope with his hands until he stops, orients his body and his gaze towards the ground where the mother is sitting (off camera) and spreads his legs, basically opening them in a way that facilitates a necessary step for a successful carry (i.e. getting his legs around mother’s body; Figure 2a). The leg spread is one of the gestures that Loto uses to elicit carries from his mother and therefore constitutes the request here (FPP). The mother immediately gets up, raising her arms to grab him and Loto extends his right arm, to reach towards the mother (Figure 2b). The mother finally grabs
him off the rope (Figure 2c) and begins carrying him around (SPP, Figure 2d). In this example, the notable element is the direction of attention of one of the two individuals (Yasa and Luiza) sitting below Loto at the time of his request. Yasa, who was examining a wooden enrichment object, looks towards Loto while he is moving along the rope. As soon as Loto produces the spread leg gesture directed towards mother (Figure 2a), Yasa looks back to the enrichment object she was looking at before paying attention to Loto (Figure 2b). In other words, Yasa monitors Loto until she sees that he is addressing somebody else and therefore does not need to react to him, but rather can resume the activity she had interrupted. Establishing a participation framework via an F-formation (through gaze direction and body orientation) means not only indicating who is addressed and part of the projectable interaction, but also who is not going to be a relevant participant in it. At times, indeed, infant bonobos request to be carried by individuals who are not their own mothers and often they succeed. Therefore monitoring whom Loto is selecting as an addressee matters not just for mother but also for the other participants in the social situation.

We have seen two examples of FPPs realized through visible gestures preceded by a selection of the addressee done via gaze. Tactile gestures allow an even clearer selection of the addressee in a group of individuals. In Example (3) Ulindi invites her infant Loto to engage in a carry by touching his back near the shoulder while walking beside him.
(3) Touch back/shoulder. Ulindi.

Figure 3a–d. Ulindi touch back/shoulder to Loto

Loto and Fimi are playing together when Yasa and Ulindi start moving beside them. While moving, Yasa looks at Fimi (Figure 3a) who perceives the look and sees mother moving. Fimi stops playing with Loto (Figure 3b) and starts following her mother Yasa (Figure 3c). Ulindi looks towards Loto just like Yasa did with Fimi (Figure 3a) but Loto is not looking at her because he is focused on Fimi. Ulindi then touches him on the back, near his left shoulder, while he is still looking at Fimi (Figure 3b). Loto turns in less than 0.2 seconds, sees Yasa, Fimi and his mother moving and immediately places his hands on Ulindi’s back (Figure 3b). Ulindi had simultaneously positioned her body to facilitate Loto’s climbing by moving her back closer to Loto and lowering it (Figure 3c). Loto climbs up his mother’s back and engages in a dorsal carry (Figure 3d).

Just like in previous examples, here we can see how mothers and infants not only monitor each other’s behavior, but specifically address their gaze and body behavior to their respective infants, who immediately respond to them. A look
back at Fimi while walking by her is Yasa’s sign that Fimi should follow her. On the other hand, Ulindi’s way of soliciting Loto to climb on her back to get carried around consists in touching his back near the shoulder while moving beside him. The time it takes both infants to begin their response is remarkably short, approximately 0.2 seconds.

The speed at which a recipient responds to a request to get carried becomes even more remarkable if the individual addressed responds before the gesture has actually reached its stroke. In Example (4), Fimi elicits a carry via a reach gesture towards her mother, while hanging from a mesh and the mother responds before Fimi’s gesture is fully articulated.

(4) Reach. Fimi.

In Figure 4a, Fimi is hanging on the mesh looking downward while the mother monitors her from behind. Fimi suddenly turns her head towards the mother and stretches her left hand towards her mother while looking towards her (therefore selecting her mother as the addressed recipient, both through gaze and the directionality of the arm reach). Before Fimi even completes the stroke of her gesture, however, the mother begins moving her right arm towards Fimi (Figure 4b). By the time Fimi reaches completely towards the mother, the
mother places her right hand under Fimi’s armpit to grab her (Figure 4c). The mother then pulls Fimi towards her venter so that she can grab onto her and initiate the carry (Figure 4d). In this example then we see that a mother can recognize what the infant is trying to communicate before the articulation of the infant’s gesture is brought to completion and the mother can therefore anticipate the production of the response. In doing so, the mother displays the capacity to project what the infant is trying to do while the signal is articulated, displaying skills (action projection and action recognition) that we know are reliably deployed by humans when they talk to each other.

Examples 1–4 are representative instances of a single gesture produced after having monitored the other participant’s attention and after having specifically addressed that individual. However, one of the advantages of using gestures is that they can be combined and produced simultaneously (if articulated via different body parts). In Example (5), Fimi, who has both reach and leg spread in her gestural repertoire, combines them in order to elicit a carry.

(5) Reach + leg spread. Fimi.

In Figure 5a, Fimi is standing on a branch close to another adult individual, towards whom she is looking, while Yasa is standing behind Fimi’s back. While the mother moves closer to Fimi, Fimi turns her head towards her mother and they look at each other (Figure 5b). After having looked at each other, and thus, after Fimi has selected her mother as her addressee and mother has made herself available as such, Fimi begins swinging on a rope. While facing her mother, Fimi stretches her right arm in a reach gesture and spreads her legs in a leg spread gesture, (Figure 5c). It is the combination of these gestures that constitutes the FPP. The mother recognizes what Fimi is requesting, and in less than 0.2 seconds moves to place her right hand behind F’s back (Figure 5d) and pulls the infant closer to her to begin the carry (Figure 5e).
As in the previous examples, this instance demonstrates the importance of the infant selecting the addressee and monitoring for her attention before producing a sequence initiating action. Further, it illustrates how gestures used in the service of similar actions (in this case, trying to elicit a carry from the mother), but articulated with different body parts, can be deployed simultaneously and obtain an appropriate and timely response by the mother. It is an empirical question what the function of producing simultaneously multiple signals with the same social function may be (i.e. whether it is a way of securing comprehension via redundancy or a way of communicating different nuances of meaning, for example, urgency).

7. Expanded sequences and pursuits

The previous examples have shown carries that were successfully accomplished after one individual had clearly selected the other as the addressed recipient of a single first signal (Examples 1–4) or a simultaneous combination of two signals (Example 5). The timing between the initiating action (FPP) and the responsive behavior (SPP) has been shown to be remarkably fast. If this is the timing of a ‘preferred’ response, then we can ask: what happens when the mother does not produce the expected or ‘preferred’ response? Delay in response can be related to a range of contingencies (e.g. because the mother is momentarily distracted or unable to proceed with a carry), and apes cannot say ‘no’ or otherwise explicitly indicate that the request will be granted, but later, so how does a bonobo know whether the recipient will respond or not? Just because the addressed recipient does not respond immediately, it does not mean that she will not respond at all. So how long do bonobos wait for an appropriate response to a carry request and what can they do if the response is not immediately forthcoming or is not the one they were aiming for?
In the following examples, I show pursuits by bonobos and a possible account for the gestures/signals that they decide to deploy, both when they first initiate their ‘request’ and when they pursue a delayed or missing response. In Example (6) the infant can see that the mother has perceived and possibly understood her ‘request’, but for reasons that will be accounted for, the mother delays approaching her and therefore engaging in a carry.

(6) Multiple pursuits. Fimi.

Figure 6a–b. Fimi arm raise + silent pout to Yasa

Fimi is hanging on a mesh not far from a male adult, and the mother is standing on two ropes close by. The mother is carefully monitoring the adult male, who is looking at Fimi, while Fimi looks towards her mother (Figure 6a) before producing the FPP: an arm raise + a silent pout with her face (Figure 6b). Each signal had successfully been used in previous instances of interaction to elicit a carry from mother. During the infant’s simultaneous production of two signals, the mother moves her gaze from the adult male to Fimi. Fimi then holds her arm raised for 0.8 seconds while the mother looks at her, then the mother moves closer to the infant but does not produce the SPP that is being sought. Without retracting her arm, Fimi moves it to produce a reach gesture, again, orienting it towards the mother (Figure 6c). At this point, mother and infant are both facing and looking at one other. Having moved on the ropes, however, the mother is not close enough to grab the infant, but she does not move closer to the infant, even though she is looking towards Fimi. This accounts for the fact that in this spatial configuration Fimi holds her reach gesture (Figure 6c) for approximately 1 second, before lowering her arm (Figure 6d).
c. d.

Figure 6c–d. Fimi reach and retraction

With Fimi’s arm lowered, the mother moves closer to the adult male on the ropes and continues to monitor the infant. This move is performed while the male keeps looking at mother and might be a way to distract his attention from the infant. It is at this point that Fimi then produces another silent pout (Figure 6e), this time without any associated arm movements, before moving closer to the male (Figure 6f).

e. f.

Figure 6e–f. Fimi silent pout and move close to Yasa

However, Fimi’s move towards the adult male appears to be what the mother was trying to avoid since she quickly moves back on the ropes towards the infant, while producing a loud vocalization looking towards the infant. As the mother moves back on the ropes and appears to prepare to jump, Fimi then produces another reach gesture (Figure 6g), thereby producing another attempt to pursue the mother’s response (i.e. getting carried). Within 0.2 seconds of this second
reach gesture, the mother jumps towards the infant, which allows Fimi to attach to her venter and simultaneously for the mother to protect the infant from the adult male, who gets up when the mother jumps (since, by doing so, she has now placed herself in between them; Figure 6h). Finally, the mother jumps back on the ropes and carries the infant away.

In this example then, the infant’s initial FPP (arm raise + silent pout) elicits some response from the mother (she moves on the ropes towards the infant); however, it is not the precise form of response Fimi’s actions sought (i.e. getting carried). The mother’s move on the ropes could be preliminary to performing the SPP, but it is not yet the SPP, so Fimi re-does the initial FPP in another (previously successful) format. Producing this transformed gesture and sustaining it for one second accomplishes the pursuit of the initial ‘request’ and therefore makes relevant the same response from the mother. In the absence of that response, and when the infant sees her mother moving away from her (i.e. Fimi can deduce that her mother will not satisfy her request because this form of response does not allow for it), Fimi produces another pursuit – the silent pout. It is only when the mother places herself in a similar spatial configuration to the one that occurred a few seconds earlier, by facing the infant again, that the infant redoes the reach gesture (compare Figure 6c and 6g). This time the mother does provide the expected SPP by approaching the infant in a way that allows her to grab onto the mother’s body and proceed with the carry.

Example (7) shows how bonobo mothers can respond to specific interactional contingencies that delay the progression of a specific course of action.

(7) Look + Present venter. Yasa.
Yasa and Fimi are sitting on a wooden platform (Figure 7a) when Yasa gets up and moves out in front of Fimi. Yasa then stops and looks towards Fimi (Figure 7b) as we have seen her doing in Example (3). We have seen how Yasa’s look towards Fimi while walking away from her is not just a way of monitoring what she is doing but also works as an invitation to move with her in the same direction. Fimi moves a bit further on the platform but does not get off it. The mother continues looking towards Fimi for about 1 second and then moves closer to Fimi and produces a present venter gesture, i.e. she positions herself belly up to invite Fimi to climb on her venter and engage in a carry (Figure 7c). Fimi hesitantly moves down the platform and closer to her mother, but this takes almost 2 seconds. Yasa begins to shake her hips sustaining the present venter position, probably to communicate to Fimi that she should hurry up. As soon as she sees her mother shaking her hips, Fimi rapidly climbs on her (Figure 7e) and Yasa carries Fimi out in the direction she had initially moved.
towards (see Figure 7b). In this example then, we can see how the mother deals with Fimi’s resistance to following her and how a slow reaction to the present venter gesture is met with a further signal (the hip shaking) most likely meant to communicate urgency. Note that the mother waits 1 second before re-approaching Fimi and orients to the slow response by Fimi to the present venter as problematic as well.

In Example (8), Loto attempts multiple times to get the mother to follow him and to engage in a carry but ultimately fails.

(8) Failed requests. Loto.

Figure 8a–b. Loto look to Ulindi

At the beginning of this sequence, Loto, his mother Ulindi and his father Kuno are sitting in the shade during a sunny and warm summer’s day. Kuno sits behind Ulindi. Loto first moves towards his mother and holds his hand on her shoulder while looking at her with his face very close to Ulindi’s face (Figure 8a). Ulindi, however, does not look at him and Loto begins to move out in front of her, until he stops and turns to look towards her (Figure 8b). This could be considered a first request for Ulindi to follow him, but Ulindi does not move.

Figure 8c–d. Loto look + wrist bent to Ulindi
Loto waits for 0.9 seconds and then moves even further until he stops and turns again to look towards Ulindi (Figure 8c). Again, Ulindi does not move. This time Loto waits for 1.2 seconds and then produces a wrist bent gesture (Figure 8d). Loto holds this gesture for 1 second waiting for a response that does not occur and then resumes walking.

While he is walking, his mother gets up and starts following him. When Loto stops walking and turns, he can see that his mother is finally following him (Figure 8e). He reaches some trees nearby and waits for her. When she arrives close to him, Loto places his hands on her back, apparently to climb on her and engage in a dorsal carry (Figure 8f). Ulindi is exactly in the same body posture as in Figure 3c in Example (3), which in that occasion led to a dorsal carry. But the lowering of Ulindi’s back, instead of being an invitation for Loto to climb on her is simply preliminary to her sitting down in the shade, which she does while Loto still has a hand on her back (Figure 8g). Loto then stops attempting to engage in a carry and squats down beside her (Figure 8h).
In this example we observe multiple attempts by the infant to get the mother to follow him and engage in a carry, which ultimately fail. After having established who his addressee is, he walks away and each turn towards the mother can be considered not only a way to monitor whether she is following him but also an invitation to follow him (as seen in Examples 3 and 7) and therefore FPPs. In the absence of a response from mother who continues sitting down, Loto resorts to another signal: the wrist bent. He usually uses this gesture to request to be carried, which is a related activity to walking together, but not the same activity. Ulindi gets up and moves towards him, but she clearly does not intend to carry him around and indeed sits down when he tries to climb onto her. Concerning the timing of each pursuit, he waits approximately 1 second following each attempt, just like Fimi and Yasa did in Examples 6 and 7. Even in the case of pursuing more than once, Loto appears to provide his mother with approximately the same amount of time to respond, as has been claimed for ordinary conversation among humans.

Example (9) shows another case of a first request to walk together and then a multiple request to get carried, this time a successful one.

(9) Two sequences. Fimi.

Figure 9a–b. First look back from Fimi

Yasa and Fimi are sitting close to each other looking straight ahead (Figure 9a) when Fimi gets up, walks few steps in front of her mother and turns towards her (FPP, Figure 9b) to invite her to move as well. The mother responds in less than 0.2s by getting up and moving in the same direction (SPP), while Fimi moves another couple of steps (Figure 9c). Fimi then stops and raises her arms, i.e. producing a gesture to elicit a carry (first request for carry FPP, Figure 9d).
Yasa however is not looking at Fimi and insteads sits down (Figure 9e). At this point Fimi walks further away and the mother follows, until Fimi stops, turns towards her mother and simultaneously produces a silent pout facial expression and a wrist bent gesture (Figure 9f).

Mother and infant move quickly towards each other, while Fimi holds her arm raised and a silent pout on her face(Figure 9g), until they engage in a carry (Figure 9h).

In this example then, we see Fimi first successfully attempting to get her mother to walk with her (Figures 9b–c) and then initially failing to elicit a carry (Figures 9d–e) because of the mother’s distraction, but then succeeding once she deploys simultaneously two signals that can by themselves elicit a carry by mother: the wrist bent gesture and the silent pout. This example shows the difference between sequential
and sequence organization. Indeed, when Fimi looks back at her mother, her arm raised, the wrist bent and the silent pout are not all part of the same sequence, even though they occur sequentially one after the other. While the look back is a FPP that obtains the expected SPP (mother walking with Fimi), the arm raised is the FPP of another sequence (one aimed at engaging in a carry) and the wrist bent and the silent pout are pursuits of this second sequence. Note also that as this is a second attempt to elicit a carry, the silent pout might also indicate frustration or disappointment because of the previous unsuccessful attempt (see de Waal 1988).

8. **Systematicity of the findings**

Having provided a qualitative description of the timing of successful SPPs (i.e. the mother approaching the infant and engaging in a carry or vice versa), I now turn to some quantitative evidence that shows how systematic the practices outlined are for the two dyads under examination. Using ELAN, I measured the time elapsed between the completion of each infant’s or mother’s gestural stroke constituting the ‘request’ (FPP) and the first body movement by the other individual that leads to the engagement of a carry (SPP). Further, I considered the time elapsed from the moment the gaze of the recipient of the FPP was oriented towards the individual producing the FPP, to measure the degree of delay in responding only once the gesture had actually been perceived by the individual addressed. Table 2 reports that in more than 80% of instances of successful carries, both Yasa and Ulindi responded with a body movement leading to a carry in 0.2 seconds or less from the completion of their infant’s last gestural stroke. Interestingly, both infants are even more likely to respond in 0.2 seconds or less to their mother’s invitation to
get carried, as both of them do so in approximately 95% of the successful carries. In all other instances the delay in responding was approximately 1 second or less.

Table 2. Distribution of carry sequences in relation to the timing of SPPs from the stroke of FPPs

<table>
<thead>
<tr>
<th>Initiator</th>
<th>Respondent</th>
<th>SPP (within 0.2 s)</th>
<th>SPP delayed (more than 0.2s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fimi (infant)</td>
<td>Yasa (mother)</td>
<td>82% (14)</td>
<td>18% (3)</td>
</tr>
<tr>
<td>Yasa</td>
<td>Fimi</td>
<td>94% (31)</td>
<td>6% (2)</td>
</tr>
<tr>
<td>Loto (infant)</td>
<td>Ulindi (mother)</td>
<td>82% (31)</td>
<td>18% (7)</td>
</tr>
<tr>
<td>Ulindi</td>
<td>Loto</td>
<td>95% (39)</td>
<td>5% (2)</td>
</tr>
</tbody>
</table>

I then investigated how many successful carries are accomplished simply through the production of a single base FPP and how many initial ‘requests’ were pursued in the absence of a response. Table 3 shows that the vast majority of successful carries were accomplished without any pursuit. Moreover, successful requests to engage in a carry were never pursued more than 3 times in these two dyads.

Table 3. Distribution of carry sequences completed without pursuits or with 1, 2 or 3 pursuits

<table>
<thead>
<tr>
<th>Initiator</th>
<th>No pursuit</th>
<th>1 pursuit</th>
<th>2 pursuits</th>
<th>3 pursuits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fimi (infant)</td>
<td>82% (14)</td>
<td>6% (1)</td>
<td>6% (1)</td>
<td>6% (1)</td>
</tr>
<tr>
<td>Yasa (mother)</td>
<td>97% (32)</td>
<td>3% (1)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Loto (infant)</td>
<td>95% (36)</td>
<td>2.5% (1)</td>
<td>2.5% (1)</td>
<td>0</td>
</tr>
<tr>
<td>Ulindi (mother)</td>
<td>98% (40)</td>
<td>2% (1)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Concerning the choices for the design of the initial FPP, there is a large variability in terms of tactile vs. visible gestures deployment (Table 4). Fimi, Loto and Yasa used more visible gestures in their first attempt to elicit a carry, while Ulindi used more tactile gestures.

Table 4. Distribution of gesture types used in first position by each subject in successful carries

<table>
<thead>
<tr>
<th>Initiator</th>
<th>Tactile gesture</th>
<th>Visible gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fimi</td>
<td>18% (3)</td>
<td>82% (14)</td>
</tr>
<tr>
<td>Yasa</td>
<td>39% (13)</td>
<td>61% (20)</td>
</tr>
<tr>
<td>Loto</td>
<td>44% (16)</td>
<td>56% (20)</td>
</tr>
<tr>
<td>Ulindi</td>
<td>59% (24)</td>
<td>41% (17)</td>
</tr>
</tbody>
</table>
Moreover, even with this limited sample of carry events it is possible to provide even further specification concerning when different types of visible or tactile gestures would be deployed. For example, if we just consider Loto we can see that the reach gesture is always articulated when the infant is hanging from a mesh or a rope or when the infant is lying on the ground, while the wrist bent gesture is always articulated when Loto is standing on the ground. Moreover, the leg spread gesture, can only be articulated when the infant is hanging from something. Similarly, Ulindi would produce a touch back/shoulder gesture if she were already standing and moving and Loto is not on her body, while she would produce a touch head gesture if Loto were already on her body or leaning on it and if she were going to get up from a sitting position. In this latter case, she would likely be communicating that she is going to move and that he should hold onto her body.

9. Discussion

This paper shows that the sequences of gestures between mother-infant bonobo dyads are structured in a manner that strongly resembles the structure of sequences of action in human social interaction. Not only do these bonobos regularly engage in successful adjacency pair-like sequences and pursue responses when they are missing, they also do so at a speed remarkably similar to the timing of ordinary human conversation. Both mothers and infants respond to one another promptly and mostly without any further pursuit by the carry sequence initiator. This suggests that at least for this kind of activity and this kind of dyads (mother-infant) the behavior of the respondent is very much cooperative and does not require much coercion or insistence by the initiator.

The findings reported in this paper also make a number of important methodological contributions to the study of communication in great apes. First, an assessment of the timing between a request and its successful response, and between a request and its pursuit, should be preliminary to any decision concerning how much time can pass between two gestures, in order for them to be considered part of the same sequence.

Second, rather than starting from gestural forms, starting from the activities in which those gestures are produced can facilitate seeing that younger great apes are not trying out gestures to determine effectiveness, but rather are relying on specific situational contingencies to decide which signal to use first and which to use next. Both the ordering of gestures and the simultaneous production of multiple gestures (e.g. spread legs + reach gesture, wrist bent + silent pout) raise the question of whether there is a possible ranking to them. That is, while efficacy
might matter (yet may also change over time), similar gestures may convey different interactional nuances, or their deployment may have to do with other contingent factors such as spatial proximity. In chimpanzees, for example, tactile gestures often precede visual gestures when combined in a sequence (Liebal, Call & Tomasello 2004). This would suggest a preference for the deployment of tactile gestures first, at least in chimpanzees. This is clearly not the case for three out of four bonobos investigated here. It is possible that this is a bias specific to the activity of eliciting carries, but another explanation seems more likely, that would also account for the large variability among the four individuals. When deciding which gesture to articulate to elicit a carry, it is more important to take into consideration the position in the environment and posture of the sequence initiator (e.g. standing, sitting, hanging, moving) and the position in the environment and posture of the addressee (close by, distant, sitting, standing) rather than the general efficacy of a gesture over another. I have indeed reported how these factors seem to play a major role in terms of selection of a visible vs. a tactile gesture. If some gestures are ontogenetically ritualized, the probability of finding individual differences in the contingencies of gesture deployment might be higher than we think. These contingencies, if neglected, would naturally affect the likelihood of finding clear patterns in data including multiple individuals. Rather, one should focus on the details of each instance of gesture production. Similarly, repeating the same gesture or using a new one is not random. If a mother does not react at all, infants tend to produce the same gesture, whereas if the mother responds, but not in a way that addresses the infant’s sequence-initiating action, then they will produce a different gesture. However, if the same gesture fails to elicit a response more than two or three times, an infant may resort to producing a different one.

Third, it is important to emphasize how bonobos, just like humans, establish a participation framework through the use of their body and gaze orientation. By orienting towards the addressee, the sequence initiator does not simply monitor the attention of the recipient but also ostensively displays who is going to be the addressee of the communicative signal. This display is public and matters not only for the addressee from whom a response might be expected but also for the other individuals present in the same environment, as it transforms them from potential addressees to bystanders.

Overall, this paper has outlined a framework that allows for a more fine-grained analysis of primate social interaction and for an assessment of the timing of interaction. The evidence provided here shows that the capacity to engage in fast-paced, repeated, cooperative adjacency-pair structures and to pursue responses flexibly by selecting among different gestures may not be uniquely human. In particular, the timing and structure of these sequences is remarkably similar to what has been observed in humans. While they might be specific to
carry sequences, these findings suggest the possibility that the structure of human social interaction and its tempo might have a more ancient evolutionary origin than we think. Indeed, if bonobos – our closest living relatives – show the same basic interactional structures that we find in human social interaction, it is conceivable that our common ancestor, living approximately 6 million years ago, may have had the skills to develop similar orderly communicative exchanges. Such a skill – to both structure and time social interaction in a human-like manner – would in fact predate the evolution of language. Once we systematically identify the interactional skills we share with other primates, we can be better equipped to understand more fully what role language specifically played in the evolution of human sociality.

Acknowledgements

I want to thank the Wolfgang Köhler Primate Research Center for generously allowing me to access their video data for the purpose of this project and Marike Schreiber, Ronny Barr and Sylvio Tüpke for producing the drawings contained in this paper. I also want to thank Marta Halina for the many helpful discussions about gestures in primates and carries in bonobos. Finally, thanks to Tanya Romaniuk for invaluable comments on an earlier version of this paper.

Notes

1. Here I am adopting the notion of preference organization of actions outlined in Schegloff (2007: 58–96); that is, the terms preferred and dispreferred refer to a structural rather than psychological relationship between a first action and a responsive one. In particular, some responsive actions align with the action produced in first position and allow for the interactional project to progress (e.g. accepting an offer) while others do not align and in some sense halt or disallow the project’s forward trajectory (e.g. rejecting an offer).

2. There are cultural differences in terms of the average onset of a preferred response; however, in most languages investigated, “on time” responses have their onset within 0.2 seconds of the completion of the FPP.

3. ELAN is an open-source software program developed at the MPI for Psycholinguistics in Nijmegen, which is freely available online for academic use: http://www.lat-mpi.eu/tools/elan. For more information on ELAN, see Wittenburg, Brugman, Russell, Klassmann & Sloetjes (2006).

4. Current work in progress (Rossano 2012b) on communication in other great apes suggests that all great apes (not just bonobos) produce orderly adjacency pairs and pursue responses in a similar manner.
References


**Author’s address**

Federico Rossano  
Department of Developmental and Comparative Psychology  
Max Planck Institute for Evolutionary Anthropology  
Deutscher Platz 6  
04103 Leipzig  
Germany  

federico_rossano@eva.mpg.de
Author’s biography

Federico Rossano. Ph.D. (Linguistics) in 2012 from Radboud University, Nijmegen (N). Worked at the MPI for Psycholinguistics from 2004 to 2008 and as a guest researcher at the MPI for Evolutionary Anthropology from 2009 to 2011. Since 2012, he is a post-doctoral researcher in the Department of Developmental and Comparative Psychology at the MPI for Evolutionary Anthropology in Leipzig (D). His research interests include: social cognition and communication in human infants and non-human primates, conversation analysis and the role of visible behavior (gesture and gaze in particular) in face-to-face interaction.