Accepted manuscript

Current Directions in Psychological Science

Primate vocal communication and the evolution of speech

Julia Fischer[1,2,3]

[1] Cognitive Ethology Laboratory, German Primate Center, Göttingen, Germany

[2] Department of Primate Cognition, Georg-August-University of Göttingen, Göttingen, Germany

[3] Leibniz ScienceCampus Primate Cognition, Göttingen, Germany

Address:

Julia Fischer, Cognitive Ethology Laboratory, German Primate Center, Kellnerweg 4, 37077 Göttingen, Germany, Fax: +49 551 3851 372, Email: jfischer@dpz.eu

Abstract

Studies of nonhuman primate communication are often motivated by the desire to shed light on the evolution of speech, as well as the socio-cognitive processes involved in nonhuman primate communication. In contrast to human speech, the vocal repertoires of nonhuman primates are evolutionarily highly conserved. Within species-specific constraints, calls may vary in relation to the internal state of the caller and/or social experience. Receivers are able to use the signals of others to predict upcoming events or behavioral dispositions. Yet, nonhuman primates do not appear to express or comprehend communicative or informative intent. Signalers are sensitive to the relation between their own actions and receivers' responses, and thus, signaling behavior can be conceived as goal-directed. As receivers are able to integrate information from multiple sources to make adaptive decisions, this renders the system flexible and powerful. Studies that take a linguistic or a biological perspective on nonhuman primate communication need to be aware of the strengths and also the limitations of their approaches. Both benefit from a focus on the mechanisms that give rise to variation in signaling and in responses to signals.

Key words: intention; meaning; primate communication; social cognition; vocalizations

Elucidating the origin of language has been declared one of the hardest problems in science (Christiansen & Kirby, 2003). Numerous studies on the communicative abilities of nonhuman primates (hereafter: "primates") were motivated by the desire understand the evolutionary roots of language (Fitch, 2010). What have we learned about the similarities and differences between primate communication and human speech along the way? And what do we know now about the representational and socio-cognitive systems underpinning primate communication?

The human language faculty draws on several subcomponents, including the development of a conventionalized signal system to map external events, objects, or thoughts onto words (semantics), a syntax engine to construct and comprehend the hierarchical structure characterizing language, as well as the ability to attribute intentions and knowledge states to others. In addition, the language faculty utilizes more general sensory-motor processes and inferential reasoning abilities (Hauser, Chomsky, & Fitch, 2002). A key question is which specific components can be thought of as precursors to abilities that are necessary to produce language, and which ones belong to the broader abilities involved in producing and comprehending language. To give an example, the ability to parse a continuous acoustic stream into meaningful units is crucial for speech comprehension, and informative for reconstructing the evolution of the language faculty, while the observation that an organism can hear sounds is less informative, although hearing is clearly also a prerequisite. Such analyses need to look beyond superficial similarities and aim to identify whether presumed precursors are indeed the substrate from which the more specific language abilities evolved. I will here focus on the vocal communication of primates, as this appears to be particularly relevant with regard to the evolution of speech, the predominant mode of human linguistic communication (Fischer & Hage, 2019; Fitch, 2010). In the first section, I will discuss the question of meaning (semantics), and briefly touch upon syntax; in the second section, I will turn to the role of social cognition in communication.

What do primate calls mean?

How we can best conceive the meaning of primate vocalizations? The text book example is the alarm call system of vervet monkeys (Seyfarth, Cheney, & Marler, 1980). In response to their three main predator types, vervets have evolved different escape strategies and also different alarm calls. Likewise, putty-nosed monkeys (Arnold & Zuberbühler, 2006) as well as Diana monkeys (Zuberbühler, 2000) produce different types of alarm calls to different types of predators. Primate calls may also vary in relation to social context (Wheeler & Fischer, 2012). For instance, chimpanzee (Slocombe & Zuberbühler, 2005) and rhesus monkey (Gouzoules, Gouzoules, & Marler, 1984) screams vary with the power

differential of the animals involved in the conflict. While such context-specific variation was initially taken as evidence for referential communication and greeted with much enthusiasm, subsequent research showed that the mechanisms that give rise to variation in monkey alarm calls and speech are fundamentally different.

Crucially, the structure of primate vocalizations is largely innate, and auditory input is not necessary to develop the species-typical vocal repertoire. Vocal production is controlled by a limbic network that activates vocal pattern generators in the brainstem. In addition, a distinct cortical network sends projections into the limbic network, thus allowing volitional control of vocal initiation and modulation. The structure of primate vocalizations is therefore generally tied to affective states. While the animals have some control over when they use the calls, they lack the neural projections necessary for volitional control of call structure. While humans share this evolutionary more ancient system of affective vocal expressions, our species has additionally evolved the neural machinery to control vocal output, and to closely integrate voicing, articulation, and breathing patterns (reviewed in Fischer & Hage, 2019).

While the production of specific call types is broadly associated with certain affective or motivational states, there is no clear one-on-one mapping. For instance, male vervet monkeys produce barks not only when they face a leopard, but also during intergroup aggression. Female vervets produce "chutters" when they spot a snake, and also when they see another group in the area (Price et al., 2015). One idea is that similar sounding calls are loosely tied to similar motivational states, such as high arousal (in the leopard and intergroup aggression context) or mild aversion (when spotting a snake or a group in the distance). Alternatively, the different situations may evoke different motivational states, but ultimately converge on the same pattern generators. Neural recordings from the relevant brain areas in freely behaving subjects may help to resolve this issue.

Irrespective of this open issue, primate vocal communication can be conceived as 'indexical': calls are indicative of some affective state, as human laughter is tied to joy. In contrast, human speech is a conventionalized communication system, where language communities agree which utterance maps onto which activity, object, or idea. Thus, speech is symbolic, while primate vocal communication is not. Moreover, the production of some call types in response to specific situations can be understood as evolved response. For instance, infant vervet monkeys spontaneously produce aerial alarms to all sorts of objects in the air, including falling leaves. Thus, they are genetically predisposed to respond to aerial "threats" with a specific call type. With growing experience, the young monkeys learn that only specific birds of prey pose a danger, and adult vervets ultimately only give aerial alarms to Martial eagles (Seyfarth et al., 1980). The analysis of the alarm calls of

East and West African members of the genus provides further evidence that the link between detecting an aerial threat and producing an aerial alarm is hardwired (Wegdell, Hammerschmidt, & Fischer, 2019). In human speech, in contrast, different words for different things are the outcome of an obligatory learning process. Depending on the linguistic communities, a toddler effortlessly acquires not only the names of the things and events in her world, she also adopts the syntactic rules of her community (Tomasello, 2008).

Although it is now firmly established that the vocal repertoires of primates are largely innate, there is some evidence that calls given by members of different social groups show slight differences or "dialects" (Crockford, Herbinger, Vigilant, & Boesch, 2004; Watson et al., 2015). Nevertheless, the variation between calls falls within the species-typical constraints. In other words, some modification of vocal structure is possible, but it does not compare to the human capacity to imitate a whole range of sounds spontaneously. For many years, researchers have butted heads over whether or not primates show vocal learning. It has therefore been suggested to overcome this dichotomous view of vocal learning and to focus instead on the mechanisms that result in variation in relation to auditory experience (e.g., Fischer & Hammerschmidt, 2020; Wirthlin et al., 2019).

For receivers, alarm calls and other vocalizations attain meaning because they can use them to predict events in the environment. Alarm calls predict the presence of a predator, a threat grunt predicts the imminence of further aggression, and the screams of a female indicate that she has been attacked and is now recruiting support. Thus, the calls of other group members can be highly informative, particularly given the fact that they are typically individually distinct (Fischer & Price, 2017). Moreover, primates are able to pick up subtle differences in calls and place them into distinct categories (Hammerschmidt & Fischer, 2008). West African green monkeys distinguish between different variants of alarm 'chirps' given in response to snakes and leopards, respectively. When the monkeys are provided with experimentally controlled information about the presence of a predator by showing either snake or leopard models, and are subsequently exposed to congruent or incongruent alarm calls, responses vary with contextual information and the information retrieved from the alarm calls (Price & Fischer, 2014).

How do primates acquire the meaning of calls and other sounds in the environment? A handful of studies have investigated the development of infant monkey responses to different call types using playback experiments. At about six months of age, vervet monkey infants respond in an adult-like fashion to different alarm calls. Similarly, baboon infants began to respond differentially to alarm barks versus contact barks at an age of six month. In contrast, Barbary macaque infants responded more strongly to their mother's calls compared to other females' calls at an age of ten weeks, while they failed to respond at all

at an age of four weeks. Taken together, variation in infant responses to their conspecific vocalizations appears to be affected by call type, variation in exposure and opportunity to learn, relevance and the general developmental stage (reviewed in Hammerschmidt & Fischer, 2008).

We recently probed how rapidly adult monkeys would learn the meaning of a novel sound. We first presented a drone to West African Green monkeys, *Chlorocebus sabaeus* (Wegdell et al., 2019). Once the monkeys spotted the drone, they gave alarm calls that sounded just like the "eagle alarms" of their East African congeners, and fled into bushes. A few days after presenting the drone, we played back the sound of the drone, to test whether the animals had associated the sound with an aerial threat. Even monkeys that had been exposed to the drone only once became vigilant and scanned the sky (Wegdell et al., 2019). Rapid auditory learning is thus possible in the wild, despite the fact that monkeys in operant conditioning tasks often struggle when auditory stimuli are used as conditioned stimuli. At the same time, the structure of the calls is evolutionarily highly conserved, as members of two species in the genus *Chlorocebus* separated by around 2 million years of evolution produced the same types of alarm calls (Wegdell et al., 2019). In summary, much of the sophistication in primate communication can be attributed to the receivers (Seyfarth & Cheney, 2003). The animals are able to integrate contextual information with information retrieved from conspecific utterances to make adaptive decisions, and they are able to rapidly learn what novel sounds predict.

Regarding the syntactical abilities of primates, a wide range of studies identified combinatorial patterns, but these are not equivalent to the rule-based compositionality of human syntax. Primates are able to process simple regularities in sequences (artificial grammars), but are vastly outperformed by humans in this regard. Current research aims at uncovering whether primates possess a "mental deep structure", which they use to represent the structure between agents, actions, and objects (Zuberbühler, 2019).

Key facets of primate vocal communication

- The structure of vocalizations is largely innate
- Modification of call structure is typically only possible within tight species-specific constraints
- There is no evidence for conventionalized symbolic communication or syntactical structures
- Receivers are able to integrate information from different sources to make adaptive decisions
- Learning plays an important role in call comprehension and the development of the appropriate responses

What is the role of social cognition in primate communication?

In human speech, linguistic meaning not only depends on the relationship between a word and what it stands for, but also on the fact that signaler and the receiver consider each other's state of mind when communicating (Moore, 2016; Scott-Phillips, 2015). In humans, the same utterance can take several different meanings, depending on the constellation between the speaker and the hearer. For instance, if I tell you "the sun is shining", I am not just uttering the observation that the sun is shining, but I also let you know that I noted this, and that I want you to note this too, for instance, because we might go out for a walk. Linguistic pragmatics encompasses the study of such intended and inferred meaning (Fitch, 2010). A second form of inference, namely "contextual pragmatics", considers the fact that receivers may use contextual cues to disambiguate the meaning of signals and to guide their decision-making in response to calls (Wheeler & Fischer, 2012). The inferential processes involved in contextual pragmatics do not appear to differ substantially between nonhuman primates and humans, while linguistic pragmatics requires an understanding of someone else's intention, knowledge, and beliefs. In its sophisticated version, linguistic pragmatics appears to be unique to human communication.

Progress in the discussion about primate intentional communication has been hampered by a lack of distinction between first-order and second-order intentionality. According to Dennett, "a first-order intentional system has beliefs and desires (etc.) but no beliefs and desires about beliefs and desires […] A second-order intentional system is more sophisticated; it has beliefs and desires […] about beliefs and desires (and other intentional states)" (Dennett, 1983, p. 345). Primate vocal communication can be well explained by assuming that callers are a first-order intentional system: the sender has a certain

knowledge state ('belief'), and the intent ('desire') to alter the behavior of others. It is not necessary to invoke higher-order intentionality.

Primate vocal behavior can be conceived as goal-directed. The signalers' goal is to evoke specific responses in receivers, and they are sensitive to whether their action (the vocal behavior) has had the desired effect. As signals are often directed, one may be inclined to presuppose intent to inform, but this is not necessary. A more parsimonious idea is that signals provide information about the signaler's motivation to engage in a specific behavior, such as play, attack, or mate. Receivers in turn learn the statistical regularities of signaling behavior and are able to consider further information for adaptive decision-making.

Building on the seminal work by Grice (1969) and Sperber (2000), Scott-Phillips (2015) argued that a decisive step in language evolution was the development of 'ostensive' communication, in which signals are produced with communicative and informative intent, and both the communicative and the informative intent are recognized. Imagine a person in a café. She first makes eye contact with the waiter, and then tilts her empty cup "in a particular, *ostensive,* way" (p. 9, italics in the original) to indicate that the cup is empty and that she desires a refill. The client assumes that the waiter understands both her communicative and informative intent. It may also be possible, however, that the client is entirely ignorant of the mental state of the waiter, and has simply learnt that a refill will be served when she tilts the cup, but only if she previously had made eye contact with the waiter. From simply observing the behavior of clients in a café, or monkeys in a forest, it is difficult to distinguish between low-level and high-level explanations. Accordingly, the issue continues to be debated (see, for instance, Moore (2017) for a proposition of a minimal Gricean model of animal communication).

Two studies in chimpanzees explicitly tested whether subjects take the knowledge states of other group members into account when they signal in response to a hidden danger only they are aware of (Crockford, Wittig, Mundry, & Zuberbühler, 2012; Schel, Townsend, Machanda, Zuberbühler, & Slocombe, 2013). The researcher placed a model of a snake near the traveling path of the animals. The question was whether a subject that was traveling ahead of the others would produce 'alert hoos' when ignorant party members arrived. Although these studies provided some evidence that the chimpanzees altered their signaling behavior in relation to the knowledge states of the others, it remains difficult to disentangle whether signalers are indeed attributing knowledge (or ignorance) or whether they are simply monitoring the others' behavior. If the latter leads to the same behavioral outcome at the side of the signaler, then there may be no need to represent what others know or believe.

A further study probed the communication in a cooperative context in the lab (Melis & Tomasello, 2019). Chimpanzees had to point to one of two boxes in which tools were hidden; these tools were necessary to access a food reward. The 'communicators' indeed indicated to conspecific 'receivers' where the tools were by moving towards and pointing at the correct box, but the salience of the box containing the tool may have affected the response. It would have been interesting to know whether informative communication would also happen in a more abstract setting, where the communicator could provide information from a distance, for instance by pushing one of two buttons that turn on lights indicating where the tools are, and without receiving immediate benefits.

Key facets of the role of social cognition in primate communication

- Primate signalers can be conceived as first-order intentional systems that operate to alter the behavior of others
- Signaling can be conceived as goal-directed
- Only few studies found some evidence that callers take the knowledge state of receivers into account
- The vast majority of instances of communication studied in primates do not provide evidence for such second-order intentionality

Conclusions and future perspectives

Call production in primates is fundamentally different from human effortless acquisition of thousands of new words early in ontogeny, and the life-long ability to acquire new words and languages (albeit with an accent). The commonality between nonhuman primates and humans rather appears to lie in the ability to process sounds and to rapidly attach meaning to new sounds. One key to understanding the evolution of speech therefore appears to lie at the side of changes in the neural control during vocal production rather than comprehension. Yet, an exquisite control over call production is only part of the answer. Another crucial aspect is the motivation to communicate about the world. According to Tomasello, "shared intentionality", the motivation to engage in social and mental coordination is a uniquely human trait. It does not only provide the foundation for linguistic communication, but also for the development of a sophisticated "Theory of Mind", that is, the ability to ascribe beliefs and knowledge to others, and to distinguish one's own mental state from that of others (Tomasello, 2008).

Nevertheless, the search for evidence for human-like linguistic or socio-cognitive abilities in primates continues. If some exceptional ability is found, researchers have to ask themselves whether they see this ability in a wider set of circumstances. For instance, if a monkey truly has the ability to label an object, why would it restrict it to leopards and eagles? If an ape routinely represents the knowledge states of others, why does this not become apparent across a broad range of situations, for instance, when a friend or foe is approaching from behind? And if a given exceptional ability is confined to a very specific set of (experimental) circumstances, then what does this tell us about the selective pressures that have brought this ability about?

Finally, we need to be aware how different research programs shape the questions and ultimately the conclusions we draw. Behavioral biologists have investigated the ontogeny, the proximate mechanisms including the neural and hormonal underpinnings of signaling, the function and the evolution of signals; in short, they have adopted a "Tinbergian" approach (Tinbergen, 1963). These studies are well suited to explain communication systems as such, but they may be silent with regard to the evolution of speech. In this realm, linguistic concepts have been more influential, but they come with a drawback: since linguistic concepts had initially been developed to describe and analyze derived and extremely specialized communication system, they miss what is ecologically and perhaps evolutionarily relevant for the animals themselves. Minding the implications of choosing one research program over another may help to place one's results into context.

Acknowledgements

I thank Kurt Hammerschmidt for discussion.

References

Arnold, K., & Zuberbühler, K. (2006). The alarm-calling system of adult male putty-nosed monkeys, *Cercopithecus nictitans martini*. *Animal Behaviour*, *72*, 643–653. doi:10.1016/j.anbehav.2005.11.017

Christiansen, M. H., & Kirby, S. (2003). The hardest problem in science? In M. H. Christiansen & S. Kirby (Eds.), *Language Evolution* (pp. 1–15). Oxford University Press. doi:10.1093/acprof:oso/9780199244843.003.0001

Crockford, C., Herbinger, I., Vigilant, L., & Boesch, C. (2004). Wild chimpanzees produce group-specific calls: A case for vocal learning? *Ethology*, *110*, 221–243. doi:10.1111/j.1439-0310.2004.00968.x

Crockford, C., Wittig, R. M., Mundry, R., & Zuberbühler, K. (2012). Wild chimpanzees inform ignorant group members of danger. *Current Biology*, *22*, 142–146. doi:10.1016/j.cub.2011.11.053

Dennett, D. C. (1983). Intentional systems in cognitive ethology: The "Panglossian paradigm" defended. *Behavioral and Brain Sciences*, *6*, 343–355. doi:10.1017/S0140525X00016393

Fischer, J., & Hage, S. R. (2019). Primate vocalization as a model for human speech: Scopes and limits. In P. Hagoort (Ed.), *Human Language: From Genes and Brains to Behavior* (pp. 639–656). Cambridge MA: MIT Press.

Fischer, J., & Hammerschmidt, K. (2020). Towards a new taxonomy of primate vocal production learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*, 20190045. doi:10.1098/rstb.2019.0045

Fischer, J., & Price, T. (2017). Meaning, intention, and inference in primate vocal communication. *Neuroscience & Biobehavioral Reviews*, *82*, 22–31. doi:10.1016/j.neubiorev.2016.10.014

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press.

Gouzoules, S., Gouzoules, H., & Marler, P. (1984). Rhesus monkey (*Macaca mulatta*) screams: Representational signalling in the recruitment of agonistic aid. *Animal Behaviour*, *32*, 182–193. doi:10.1016/S0003-3472(84)80336-X

Hammerschmidt, K., & Fischer, J. (2008). Constraints in primate vocal production. In D. K. Oller & U. Griebel (Eds.), *Evolution of Communicative Flexibility: Complexity, Creativity, and Adaptability in Human and Animal Communication* (pp. 93–119). Cambridge MA: MIT Press. doi:10.7551/mitpress/9780262151214.001.0001

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it,

who has it, and how did it evolve? *Science*, *298*, 1569–1579. doi:10.1126/science.298.5598.1569

Melis, A. P., & Tomasello, M. (2019). Chimpanzees (*Pan troglodytes*) coordinate by communicating in a collaborative problem-solving task. *Proceedings of the Royal Society B: Biological Sciences*, *286*, 20190408. doi:10.1098/rspb.2019.0408

Moore, R. (2016). Meaning and ostension in great ape gestural communication. *Animal Cognition*, *19*, 223–231. doi:10.1007/s10071-015-0905-x

Moore, R. (2017). Convergent minds: Ostension, inference and Grice's third clause. *Interface Focus*. doi:10.1098/rsfs.2016.0107

Price, T., & Fischer, J. (2014). Meaning attribution in the West African green monkey: Influence of call type and context. *Animal Cognition*, *17*, 277–286. doi:10.1007/s10071-013-0660-9

Price, T., Wadewitz, P., Cheney, D. L., Seyfarth, R. M., Hammerschmidt, K., & Fischer, J. (2015). Vervets revisited: A quantitative analysis of alarm call structure and context specificity. *Scientific Reports*, *5*, 13220. doi:10.1038/srep13220

Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., & Slocombe, K. E. (2013). Chimpanzee alarm call production meets key criteria for intentionality. *PLoS ONE*, *8*, e76674. doi:10.1371/journal.pone.0076674

Scott-Phillips, T. C. (2015). *Speaking Our Minds*. London & New York: Palgrave Macmillan.

Seyfarth, R. M., & Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annual Review of Psychology*, *54*, 145–173. doi:10.1146/annurev.psych.54.101601.145121

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, *210*, 801–803. doi:10.1126/science.7433999

Slocombe, K. E., & Zuberbühler, K. (2005). Agonistic screams in wild chimpanzees (*Pan troglodytes schweinfurthii*) vary as a function of social role. *Journal of Comparative Psychology*, *119*, 67–77. Retrieved from isi:000227365100008

Tinbergen, N. (1963). On aims and methods in ethology. *Zeitschrift für Tierpsychologie*, *20*, 410–433.

Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.

Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., … Slocombe, K. E. (2015). Vocal learning in the functionally referential food grunts of

chimpanzees. *Current Biology*, *25*, 495–499. doi:10.1016/j.cub.2014.12.032

Wegdell, F., Hammerschmidt, K., & Fischer, J. (2019). Conserved alarm calls but rapid auditory learning in monkey responses to novel flying objects. *Nature Ecology & Evolution*, *3*, 1039–1042. doi:10.1038/s41559-019-0903-5

Wheeler, B. C., & Fischer, J. (2012). Functionally referential signals: A promising paradigm whose time has passed. *Evolutionary Anthropology*, *21*, 195–205. doi:10.1002/evan.21319

Wirthlin, M., Chang, E. F., Knörnschild, M., Krubitzer, L. A., Mello, C. V, Miller, C. T., … Yartsev, M. M. (2019). A Modular Approach to Vocal Learning: Disentangling the Diversity of a Complex Behavioral Trait. *Neuron*, *104*, 87–99. doi:10.1016/j.neuron.2019.09.036

Zuberbühler, K. (2000). Referential labelling in Diana monkeys. *Animal Behaviour*, *59*, 917–927. doi:10.1006/anbe.1999.1317

Zuberbühler, K. (2019). Evolutionary roads to syntax. *Animal Behaviour*, *151*, 259–265. doi:10.1016/j.anbehav.2019.03.006

Recommended Readings

Fischer, J. (2017). *Monkeytalk: Inside the worlds and minds of primates*. Chicago: University of Chicago Press.

- This book provides an introduction to monkey social behavior, cognition, and communication and is addressed to a broader audience

Fischer, J., & Price, T. (2017). Meaning, intention, and inference in primate vocal communication. Neuroscience & Biobehavioral Reviews, 82, 22–31. doi:10.1016/j.neubiorev.2016.10.014

- This paper provides a more elaborate discussion of the question of the attribution of meaning and the socio-cognitive processes underpinning nonhuman primate communication.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. Science, 210, 801–803. doi:10.1126/science.7433999

- This study is certainly the single most influential paper in this field, paving the way for an enormously productive research program addressing primate communication and social cognition

-