Transition to language: From agent perception to event representation

Klaus Zuberbühler\textsuperscript{1,2} | Balthasar Bickel\textsuperscript{3,4}

\textsuperscript{1}Institute of Biology, University of Neuchatel, Neuchatel
\textsuperscript{2}School of Psychology and Neuroscience, University of St Andrews, St Andrews
\textsuperscript{3}Department of Comparative Language Science, University of Zurich, Zurich, Switzerland
\textsuperscript{4}Center for the Interdisciplinary Study of Language Evolution, University of Zurich, Zurich, Switzerland

Correspondence
Klaus Zuberbühler, Institute of Biology, University of Neuchatel, Rue Emile Argand 11, CH-2000 Neuchatel, Switzerland.
Email: klaus.zuberbuehler@unine.ch

Funding information
We are grateful to the Swiss National Science Foundation (grant numbers 310030\_185324 and 100015\_182845) with further funding from the NCCR Evolving Language (SNSF 51NF40\_180888).

Edited by: Juan-Carlos Gómez, Editor

Abstract
Spoken language, as we have it, requires specific capacities—at its most basic advanced vocal control and complex social cognition. In humans, vocal control is the basis for speech, achieved through coordinated interactions of larynx activity and rapid changes in vocal tract configurations. Most likely, speech evolved in response to early humans perceiving reality in increasingly complex ways, to the effect that primate-like signaling became unsustainable as a sole communication device. However, in what ways did and do humans see the world in more complex ways compared to other species? Although animal signals can refer to external events, in contrast to humans, they usually refer to the agents only, sometimes in compositional ways, but never together with patients. It may be difficult for animals to comprehend events as part of larger social scripts, with antecedent causes and future consequences, which are more typically tie the patient into the event. Human brain enlargement over the last million years probably has provided the cognitive resources to represent social interactions as part of bigger social scripts, which enabled humans to go beyond an agent-focus to refer to agent–patient relations, the likely foundation for the evolution of grammar.

This article is categorized under:
- Cognitive Biology > Evolutionary Roots of Cognition
- Linguistics > Evolution of Language
- Psychology > Comparative

KEYWORDS
- event perception
- evolution of grammar
- script theory
- theory of mind
- vocal learning

1 | HALLMARKS OF LANGUAGE

Humans are unique in a number of ways but perhaps most strikingly in their capacity for language. No other animal species has a communication system even remotely comparable, a fact that has puzzled scholars for centuries (Darwin, 1871). How could a complex faculty, such as language, have evolved during the relatively short evolutionary history of our species? One way of studying the problem is to decompose language into core properties and investigate
the associated capacities separately (Hurford, 2007). To this end, studying the communication systems of primates and other animals has been particularly enlightening. The approach is based on the premise that evolutionary adaptations are often modifications of preexisting functions rather than true innovations. While this approach has revealed continuities between animal and human communication, particularly regarding the attribution of meaning, it has also highlighted ways by which humans deviate substantially from what is normally observed in primate communication, specifically vocal control and social cognition (W. T. Fitch & Zuberbühler, 2013).

2 | EVOLUTION OF VOCAL CONTROL

By default, language is strongly tied to the acoustic mode. From early on infants begin to playfully babble with sounds, a behavior not observed in other primate species (Oller et al., 2013). For instance, chimpanzee infants remain mainly silent during their first years (Laporte & Zuberbuehler, 2011), babbling is a rare exception in primate vocal development (Elowson et al., 1998) and as adults primates continue to have limited vocal control and no real capacity for vocal learning, even with substantial training (Cheney & Seyfarth, 2005). It has also been noted that great apes are not interested in imitating human speech although, in other contexts, they like copying human behavior. Hayes and Hayes (1951) write about their home-raised chimpanzee “Viki”:

... Just as the human child copies its parents’ routine chores, so Viki dusts, washes dishes, sharpens pencils, saws, hammers ... [...]. On the other hand, she is less vocal: while the human child commonly keeps up an almost continual stream of chatter—with or without meaning, Viki is silent.

Why are such seemingly trivial vocal imitation tasks near impossible for primates? One popular explanation has been to invoke anatomical differences of the vocal tract, particularly the permanently lowered human larynx (Lieberman, 2012). But this view is no longer supported by current research. The primate vocal tract, as it has been put, is perfectly speech-ready, despite the fact that speech has only evolved in humans (W. Fitch et al., 2016). Also relevant is that great apes are capable of controlling the supra-laryngeal vocal tract to produce various voiceless signals, such as clicks, smacks, raspberries, kiss-sounds, and whistles (Lameira et al., 2014), suggesting that key articulatory movements have evolved before speech. Furness (1916) wrote of his home-raised orangutan:

The orang in one respect does use the lips, to make a sound indicating warning or apprehension; this sound is made with the lips pursed up and the air sucked through them ... [...] My oldest orang would make this sound on command (I had merely to say “What is the funny sound you make when you are frightened?”).

Producing sustained airflow to activate the vocal folds, however, appears to be much harder for nonhuman primates. Related, it is difficult to train primates to vocalize on command, suggesting that control over the larynx and respiratory muscles may have evolved more recently. On this, Hayes and Hayes (1951) write:

The first step was aimed at teaching her merely to vocalize on command, in order to obtain a reward. [...] The task was surprisingly difficult. Although she seemed to learn what was required quickly, she had serious trouble with the motor skill of voluntary vocalization.

This is remarkable because vocal learning appears to evolve relatively easily across the animal kingdom (Vernes et al., 2021) although usually for functions related to reproduction or social cohesion. Examples are birdsong (Doupe & Kuhl, 1999) and or the song of cetaceans (Janik, 2009), dolphin whistle imitations (Janik, 2000) or vocal convergence in primate social groups, usually as a means to signal social closeness (Ruch et al., 2018). Human speech, on the other hand, does not just function in sexual rivalry or as social markers; it is a semantic device allowing speakers to refer to reality the way they perceive or remember it.

One hypothesis is that the transition to vocal control in humans occurred in the context of cooperative breeding (Hrdy, 2009; Zuberbühler, 2011). Humans are unusual in the amount of childcare they provide, both in traditional hunter-gatherer and in modern societies. Often, this involves unrelated individuals, and this may especially be challenging for infants. Advanced vocal control, beyond a basic primate-like call repertoire, may have evolved to help infants to secure care from older individuals who often do not have a genetic interest to do so. In particular, natural selection may
have favored vocal behavior in early human infants that facilitated attention getting and social bonding, especially with unrelated individuals. Babbling may be particularly important (Darwin, 1871), if it helps human infants to be noted by caregivers and to facilitate social bonding with them. Interestingly, babbling is generally perceived as “pleasant,” suggesting that natural selection may have benefitted from a preexisting receiver bias. The hypothesis that babbling elicits infant care more efficiently than other vocal behavior requires further testing, ideally across different primate species.

Another hypothesis is that vocal control evolved in response to the demands of higher social cognition. Human evolution is characterized by a considerable brain enlargement starting around 1 mya (Ponce de León et al., 2021), which must have led to more complex representational abilities, especially in the social domain (Dunbar & Shultz, 2007). As humans were able to mentally represent social information in more complex ways, this may have created selection pressure on the communication system. Increased vocal control may have been an evolutionarily easy solution, endowing early humans with an ability to create a wider range of utterances to refer to the world than with a rigid signal repertoire. The hypothesis that vocal control may have evolutionarily followed advanced representational abilities is in line with research on sign language, which has shown that language does not require speech.

### 3 FROM COMBINATORIAL TO COMPOSITIONAL

Animal communication can be meaningful at various levels. First, if an individual consistently produces the same signal to a specific event, the signal becomes a predictor, a relation animals understand quickly by associative learning. But meaning is sometimes conveyed in more complicated ways, for example, via the context of signal production. For example, Guinea fowls in Tai Forest alarm call to both humans and leopards. Diana monkeys appear to be aware of this as they respond differently to the birds’ alarm calls depending on whether they suspect humans or leopards to be the cause of the calls (Zuberbühler, 2000). Extracting meaning from taking context into account may in fact be the default mechanism in animal communication (R. Seyfarth & Cheney, 2018; Smith, 1977). Here, there are interesting links to pragmatic abilities in humans, although these go beyond what is normally seen in animal communication, mainly due to the fact that they involve generalised reasoning about intentions (Levinson, 2000). Although great apes (and probably other animals) can assess others’ intentions during communication (e.g., Genty et al., 2015), they do not appear to make active use of their own intentions during signal production to generate pragmatic meaning.

Finally, meaning in animal communication can also be transmitted by signal combinations. Examples are unordered sequences where meaning resides in the distribution of bigrams (e.g., duplications): in titi monkeys, listeners react to the proportion of one type of bigram relative to others, allowing them to infer the type and location of danger (Berthet et al., 2019). Similarly, bonobos produce call sequences with bark or peep bigrams to preferred foods and peep-yelp and yelp bigrams to non-preferred foods (Zuberbühler, 2020). In playback experiments listener responded in appropriate ways, suggesting they managed to extract meaning from these sequences (Clay & Zuberbühler, 2011).

Sometimes, signal sequences have identifiable orders (i.e., permutations). This has been reported, for example, in Diana monkey contact calls (Coye et al., 2018), Campbell’s monkey alarm call suffixations, Campbell’s monkey boom/alarm combinations (Ouattara et al., 2009) or putty-nosed monkey pyow-hack combinations (Arnold & Zuberbühler, 2006). In all cases, there is corresponding experimental evidence that listeners attend to the changes in meaning induced by the permutations (Arnold & Zuberbühler, 2008; Coye et al., 2015, 2016; Zuberbühler, 2002). Great apes also produce permutations (Crockford & Boesch, 2005; Fedurek et al., 2016; Leroux et al., 2021) but it is currently unclear how the meaning of calls and sequences interact.

Per definition, compositionality requires that the meaning of a complex expression depends on the meaning of its parts and the way they are combined, that is, the meanings of the composing units and the way they related to one another (Johnson, 2021). Campbell’s monkey suffixation is an interesting example, because the -oo modulation, given to non-dangerous eagle or leopard-related situations, operates like an affix in human languages (e.g., the plural ending -s for English nouns) (Ouattara et al., 2009). Affixes are meaningful but they cannot occur on their own and bear a specific meaning relation to their host. There are also good candidates for compositionality from other groups of animals, especially in birds (Engesser et al., 2016; Suzuki et al., 2016), although there are alternative explanations (Arnold & Zuberbühler, 2012; Schlenker et al., 2016) and animal compositionality might be restricted to simple “and” relations (Townsend et al., 2018).
One key area where human languages exploit compositionality is in event coding. Many sentences consist of arguments that describe how agents as causal forces bring about states of affairs to patients. There is a strong trend in languages to have explicit grammatical signals that allow listeners to distinguish the agent from a patient and the nature of their interaction. English uses word order, other languages use special markers. Across languages, agents are perceived as privileged insofar as they exert physical or psychological power over patients or states of affairs (Sauppe & Flecken, 2021; Dowty, 1991).

Do animals perceive the world in the same way, that is, as causally structured agent and patient interactions? One source of information comes from ape language studies (Tomasello, 2017). Here, apes have demonstrated that they can understand their human caretakers to an astonishing degree, including instructions that contain agent–patient relations (Savage-Rumbaugh et al., 1993). At the same time, however, there was no compelling evidence that any of the language-trained apes were capable of constructing basic sentences with agent–patient interactions. Research on natural animal communication shows the same pattern. On the one hand, playback experiments have shown that monkeys and apes can make basic inferences about who-does-what-to-whom, by simply listening to call exchanges between two individuals (Bergman et al., 2003; Slocombe et al., 2010). But again, in terms of production, there are no good examples from any species indicating that animals can refer to the agent–patient relations of an event (Wilson et al., 2022; Zuberbühler, 2021). Hence, the tentative current conclusion is that animals understand social interactions in terms of agency, but are unable to convey this in their signal production, in line with previously reported signaler–receiver differences in animal communication (R. M. Seyfarth & Cheney, 2003). The currently known cases of compositionality in animal communication refer to either the agent, for instance by producing acoustically distinct suffixed or unsuffixed alarm calls that specify not only the predator type but also its current level of threat (Ouattara et al., 2009), or to desired responses by recipients (Engesser et al., 2016) with no reference to patients. Animal vocalisations are often self-reports which denote agents or patients, but never both (e.g., “I am digging”; “I am moving”: Jansen et al., 2012). Nonhuman primates, as it has been eloquently put, are “dendrophobic” (unable to mentally represent or manipulate tree structures) whereas human infants gradually become dendrophilic during their linguistic development (Trueswell, 2017).

Why are humans prone to integrate patients with agents into hierarchical structures, at least in their communication? Natural events are embedded in time, with past histories and future consequences. If an animal witnesses an event (e.g., A bites B) then this is most likely a consequence of a prior event (e.g., B threatened A’s infant) and will likely have future consequences (e.g., B avoids A). Human event cognition may be more complex insofar as events are not perceived as stand-alone experiences, but as embedded in time with a past and future. Perceiving events in time, according to this hypothesis, gives patients shadow agency as active participants (either in the past or future); they participate in larger scripts of how social interactions unfold with constantly changing agency (Fillmore, 1982). Humans, when witnessing an agent acting upon a patient, may already see the patient as the future agent. If this is true then grammar may only evolve in minds that can represent the world beyond the here-and-now, with its natural focus on agents. Related to this and in analogy to the main sentence functions of human language (i.e., declarative, interrogative, exclamative, and imperative), animal communication appears to be restricted to the exclamative and imperative, that is, functions in which the attention is on the present state and the agent.

Recently, we have argued that humans (and animals to various degrees) solve social problems by activating behavior scripts, accumulated with ontogenetic experience and phylogenetic predisposition (Taylor et al., under review; Zuberbühler, 2022). Behavior scripts are patterns of how interactions typically unfold in reality and are stored in memory as such. Behavior scripts have a temporal dimension with entry points, allowing subjects to recognize and categorize ongoing scripts from incomplete event input, and make predictions about what will happen next. We also argued that behavior script theory can explain subjects’ performance in knowledge-ignorance paradigms (false-belief tasks), such the Sally–Anne test (human infants) (Baron-Cohen et al., 1985), the snake paradigm (wild chimpanzees) (Crockford et al., 2012) or the King Kong paradigm (captive apes) (Krupenye et al., 2016). Briefly, in the Sally–Anne test, children see two puppets interacting. The gist of the story is that Anne removes Sally’s marble in her absence. Children then have to predict where Sally, upon her return, will look for her marble. From around 4 years, children say that Sally will search where she left her marble (old location), while younger children say that Sally will search where Anne moved the marble (new location), suggesting that attributing false beliefs is part of cognitive development. However, it
is conceivable that children solve the Sally–Anne test by reference to a simple behavior script, for example, that “people return to where they have left stuff,” something they may learn from own experience or from observing others. Similar behavior-based arguments can be made for false-belief experiments with apes (Taylor et al., under review). Complex theory-of-mind based notions, such as “Anne has a false belief” may be linguistic shortcuts to refer to situations where two behavior scripts are in conflict and require a decision (Sally goes to the marble; Sally goes to where she left the marble). The fact that only 4-year olds respond correctly may simply reflect the fact that, at a younger age, children do not have enough experience to understand that current events can be explained by past events.

Indeed, scripts (or “frames”) are well-known to be the critical basis for semantics and grammar in language (Fillmore, 1982). They are an essential part of the evolutionary package that languages endow humans with. The depth of event perception, in other words, may depend on an individual’s resources and experience, both cognitively and linguistically. The key variable is how much an individual can deviate from the present and recognize events as part of larger scripts of event sequences that are logically connected and unfold according to basic laws of social behavior. The more restricted to the present, the more individuals will focus on agents. With increasing perception of time, the same agent–patient interaction will be increasingly perceived as part of larger script. Perceiving the world in this way will turn attention to the patients of ongoing events and in the corresponding signal output.

6 | CONCLUSIONS

How did human communication transition from primate-like vocal behavior to grammatically organized spoken language? The argument is that speech was made possible by granting a speech-ready vocal apparatus voluntary control over sound production, in concert with already existing articulatory control. This was likely an evolutionary response to early humans perceiving social interactions in increasingly complex ways, with corresponding demands on the communication system. This may have been in terms of perceiving events beyond their present appearance, to include a timeline of prior causes and future consequences. Adding such a temporal dimension to event perception will reduce the strong agent focus that appears to dominate animal cognition, in favor of more balanced agent–patient perception. Humans may be the only species capable of integrating reference to a patient with reference to an agent, thus removing a major obstacle in the evolution of grammar.

AUTHOR CONTRIBUTIONS

Klaus Zuberbühler: Conceptualisation (equal); funding acquisition (equal); Writing – original draft. Balthasar Bickel: Conceptualisation (equal); funding acquisition (equal); Writing – review and editing.

ACKNOWLEDGMENTS

We are grateful to members of the NCCR Evolving Language (www.evolvinglanguage.ch) for so many inspiring discussions. Open access funding provided by Universite de Neuchatel.

CONFLICT OF INTEREST

The author have declared no conflicts of interest for this article.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Klaus Zuberbühler © https://orcid.org/0000-0001-8378-088X
Balthasar Bickel © https://orcid.org/0000-0002-9087-0565

RELATED WIREs ARTICLES

Events parsing and the origins of grammar
Linguistic capacity of non-human animals

REFERENCES

Arnold, K., & Zuberbuhler, K. (2006). Semantic combinations in primate calls. Nature, 441(7091), 303 <Go to ISI>: 000237593200035.
Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. Current Biology, 18(5), R202–R203. https://doi.org/10.1016/j.cub.2008.01.040

Arnold, K., & Zuberbühler, K. (2012). Call combinations in monkeys: Compositional or idiomatic expressions? Brain and Language, 120(3), 303–309. https://doi.org/10.1016/j.bandl.2011.10.001

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic-child have a theory of mind. Cognition, 21(1), 37–46 <Go to ISI>: //A1985AVC5100002.

Bergman, T. J., Beehner, J. C., Cheney, D. L., & Seyfarth, R. M. (2003). Hierarchical classification by rank and kinship in baboons. Science, 302(5648), 1234–1236. https://doi.org/10.1126/science.1087513

Berthet, M., Mesbahi, G., Pajot, A., Casar, C., Neumann, C., & Zuberbühler, K. (2019). Titir monkeys combine alarm calls to create probabilistic meaning. Science Advances, 5(5), eaav3991. https://doi.org/10.1126/sciadv.aav3991

Cheney, D. L., & Seyfarth, R. M. (2005). Constraints and preadaptations in the earliest stages of language evolution. Linguistic Review, 22(2–4), 135–159 <Go to ISI>: //WWW00235918400003.

Clay, Z., & Zuberbühler, K. (2011). Bonobos extract meaning from call sequences. PLoS One, 6(4), e18786. https://doi.org/10.1371/journal.pone.0018786

Coye, C., Ouattara, K., Arlet, M. E., Lemasson, A., & Zuberbühler, K. (2018). Flexible use of simple and combined calls in female Campbell's monkeys. Animal Behaviour, 141, 171–181. https://doi.org/10.1016/j.anbehav.2018.05.01

Coye, C., Ouattara, K., Zuberbühler, K., & Lemasson, A. (2015). Suffixation influences receivers' behaviour in non-human primates. Proceedings of the Royal Society B—Biological Sciences, 282(1807), 50265 <Go to ISI>: //CCC000357460400008.

Coye, C., Zuberbühler, K., & Lemasson, A. (2016). Morphologically structured vocalizations in female Diana monkeys. Animal Behaviour, 115, 97–105. https://doi.org/10.1016/j.anbehav.2016.03.010

Crockford, C., & Boesch, C. (2005). Call combinations in wild chimpanzees. Behaviour, 142, 397–421 <Go to ISI>: //000230336200001.

Crockford, C., Wittig, R. M., Mundry, R., & Zuberbühler, K. (2012). Wild chimpanzees inform ignorant group members of danger. Current Biology, 22(2), 142–146. https://doi.org/10.1016/j.cub.2011.11.053

Darwin, C. (1871). The descent of man and selection in relation to sex. Murrey/Random House.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. Annual Review of Neuroscience, 22, 567–631.

Dowty, D. R. (1991). Thematic proto-roles and argument selection, 67, 547–619. Language.

Dunbar, R. I. M., & Shultz, S. (2007). Evolution in the social brain. Science, 317(5843), 1344–1347. https://doi.org/10.1126/science.1145463

Elowson, A. M., Snowdon, C. T., & Lazaro-Perea, C. (1998). ‘Babbling’ and social context in infant monkeys: Parallels to human infants. Trends in Cognitive Sciences, 2(1), 31–37 <Go to ISI>: //000078870500005.

Engesser, S., Ridley, A. R., & Townsend, S. W. (2016). Meaningful call combinations and compositional processing in the southern pied babbler. Proceedings of the National Academy of Sciences of the United States of America, 113(21), 5976–5981. https://doi.org/10.1073/pnas.1600970113

Fedurek, P., Zuberbühler, K., & Dahl, C. D. (2016). Sequential information in a great ape utterance. Scientific Reports, 6, 38226.

Fillmore, C. J. (1982). Frame semantics. Linguistics in the morning calm, 95, 111–137. Seoul: Hanshin.

Fitch, W. T., de Boer, B., Mathur, N., & Ghazanfar, A. (2016). Monkey vocal tracts are speech-ready. Current Biology, 22(2), 142–146. https://doi.org/10.1016/j.cub.2011.11.053

Furness, W. H. (1916). Observations on the mentality of chimpanzees and orang-utans. Proceedings of the American Philosophical Society, 55, 281–290.

Genty, E., Neumann, C., & Zuberbühler, K. (2015). Bonobos modify communication signals according to recipient familiarity. Scientific Reports, 5, 16442. https://doi.org/10.1038/srep16442

Hayes, K. J., & Hayes, C. (1951). The intellectual development of a home-raised chimpanzee. Proceedings of the American Philosophical Society, 95, 105–109.

Hrdy, S. B. (2009). Mothers and others: The evolutionary origins of mutual understanding. Belknap Press.

Hurford, J. R. (2007). The origins of meaning. Oxford University Press.

Janik, V. M. (2000). Whistle matching in wild bottlenose dolphins (Tursiops truncatus). Science, 289, 1355–1357.

Janik, V. M. (2009). Whale song. Current Biology, 19(3), R109–R111. https://doi.org/10.1016/j.cub.2008.11.026

Jansen, D. A. W. A. M., Cant, M. A., & Manser, M. B. (2012). Segmental concatenation of individual signatures and context cues in banded mongoose (Mungos mungo) close calls. BMC Biology, 10, 1–10.

Johnson, M. (2021). Compositionality in language. Internet Encyclopedia of Philosophy https://iep.utm.edu/composit/

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. Science, 354(6308), 110–114. https://doi.org/10.1126/science.aaa8110

Lameira, A. R., Maddieson, I., & Zuberbühler, K. (2014). Primate feedstock for the evolution of consonants. Trends in Cognitive Sciences, 18(2), 60–62. https://doi.org/10.1016/j.tics.2013.10.013

Laporte, M. N. C., & Zuberbühler, K. (2011). The development of a greeting signal in wild chimpanzees. Developmental Science, 14(5), 1220–1234. https://doi.org/10.1111/j.1467-7687.2011.01069.x
Leroux, M., Bosshard, A. B., Chandia, B., Manser, A., Zuberbühler, K., & Townsend, S. W. (2021). Chimpanzees combine pant hoots with food calls into larger structures. *Animal Behaviour*, 179, 41–50. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85110679946&doi=10.1016%2fj.anbehav.2021.06.026&partnerID=40&md5=e3fa3f93f85d68b1bd382f9f865f81cd

Levinson, S. C. (2000). *Presumptive meanings: the theory of generalised conversational implicature*. Cambridge, MA: MIT Press.

Lieberman, P. (2012). Vocal tract anatomy and the neural bases of talking. *Journal of Phonetics*, 40(4), 608–622. https://doi.org/10.1016/j.joph.2012.04.001

Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences of the United States of America*, 110(16), 6318–6323. https://doi.org/10.1073/pnas.1300337110

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell’s monkeys use affixation to alter call meaning. *PLoS One*, 4(11), e7808. https://doi.org/10.1371/journal.pone.0007808

Ponce de León, M. S., Bienvenu, T., Marom, A., Engel, S., Tafforeau, P., Warren, J. L. A., Lordkipanidze, D., Kurniawan, I., Murti, D. B., Suriyanto, R. A., Koesbardiati, T., & Zollikofer, C. P. E. (2021). The primitive brain of early homo. *Science*, 372(6538), 165. https://doi.org/10.1126/science.aaz0032

Ruch, H., Zürcher, Y., & Burkart, J. (2018). The function and mechanism of vocal accommodation in humans and other primates. *Biological Reviews*, 93(4), 1469–185X [Electronic], 996–1013. https://doi.org/10.1111/bvr.12382

Sauppe, S., & Flecken, M. (2021). Speaking for seeing: Sentence structure guides visual event apprehension. *Cognition*, 206, 104516. https://www.sciencedirect.com/science/article/pii/S0010027720303358

Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., & Brakke, K. E. (1993). *Language comprehension in ape and child*. Lawrence Erlbaum Associates, Hillsdale.

Schlenker, P., Chemla, E., Arnold, K., & Zuberbühler, K. (2016). Pyow-hack revisited: Two analyses of putty-nosed monkey alarm calls. *Lingua*, 171, 1–23. https://doi.org/10.1016/j.lingua.2015.10.002

Seyfarth, R., & Cheney, D. (2018). Pragmatic flexibility in primate vocal production. *Current Opinion in Behavioral Sciences*, 21, 56–61. https://doi.org/10.1016/j.cobeha.2018.02.005

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

Slocombe, K. E., Kaller, T., Call, J., & Zuberbuehler, K. (2010). Chimpanzees extract social information from agonistic screams. *PLoS One*, 5(7), e11473. https://doi.org/10.1371/journal.pone.0011473

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

Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nature Communications*, 7, 10986. https://doi.org/10.1038/ncomms10986

Taylor, D., Göntül, G., Alexander, C., Glock, H.-J., Clément, F., & Zuberbühler, K. (under review). Reading minds or reading scripts? De-intellectualising theory of mind. https://osf.io/qxe9z/?view_only=d9a3e3d1e6d1480f97180ae69588a4ca

Tomassello, M. (2017). *What did we learn from the ape language studies?* Oxford Scholarship. https://doi.org/10.1093/oso/9780198729511.001.0001

Townsend, S. W., Engesser, S., Stoll, S., Zuberbühler, K., & Bickel, B. (2018). Compositionality in animals and humans. *PLOS Biology*, 16(7), e2006425.

Truswell, R. (2017). Dendrophobia in bonobo comprehension of spoken English. *Mind & Language*, 32, 395–415.

Vernes, S. C., Janik, V. M., Fitch, W. T., & Slater, P. J. B. (2021). Vocal learning in animals and humans. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 376(1836), 20200234. https://doi.org/10.1098/rstb.2020.0234

Wilson, V. A. D., Zuberbühler, K., & Bickel, B. (2022). The evolutionary origins of syntax: Event cognition in nonhuman primates. *Science Advances*, in press.

Zuberbühler, K. (2000). Causal cognition in a non-human primate: Field playback experiments with Diana monkeys. *Cognition*, 76(3), 195–207. https://doi.org/10.1016/s0010-0277(00)00079-2

Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour*, 63, 293–299. https://doi.org/10.1006/anbe.2001.1914

Zuberbühler, K. (2011). Cooperative breeding and the evolution of vocal flexibility. In M. Tallerman & K. Gibson (Eds.), *Oxford handbook of language evolution*. Oxford University Press.

Zuberbühler, K. (2020). Syntax and compositionality in animal communication. *Philosophical Transactions of the Royal Society B*, 375(1789), 20190062.

Zuberbühler, K. (2021). Event parsing and the origins of grammar. *WIREs Cognitive Science*, e1587.

Zuberbühler, K. (2022). Social scripts as drivers of primate cooperation. *Int&Vertebrates*, https://doi.org/10.52732/BORX9839

---

**How to cite this article:** Zuberbühler, K., & Bickel, B. (2022). Transition to language: From agent perception to event representation. *WIREs Cognitive Science*, e1594. https://doi.org/10.1002/wcs.1594