SEQUENTIAL DYNAMICS IN THE VOCAL SYSTEM OF COMMON MARMOSETS (*CALLITHRIX JACCHUS*)

Alexandra B. Bosshard[1,2,*], Balthasar Bickel[1,2], Paola Merlo[3], Judith Burkart[2,4] and Simon W. Townsend[1,2,5]

[*]Corresponding Author: alexandra.bosshard@uzh.ch
[1]Department of Comparative Language Science, University of Zurich, Zurich,
[2]Center for the Interdisciplinary Study of Language Evolution (ISLE), University of Zurich, Zurich, Switzerland
[3]Department of Linguistics University of Geneva, Geneva, Switzerland
[4]Department of Anthropology, University of Zurich, Zurich, Switzerland
[5]Department of Psychology, University of Warwick, Coventry, UK

Over the last 20 years there has been growing interest in the combinatorial abilities of animals, namely the propensity to sequence context-specific calls (see e.g., Arnold & Zuberbühler, 2006; Berthet et al., 2019; Coye, Ouattara, Zuberbühler, & Lemasson, 2015; Coye, Zuberbühler, & Lemasson, 2016), given the light such data can shed on the evolutionary progression of our own combinatorial communication system (Leroux & Townsend, 2020; Townsend, Engesser, Stoll, Zuberbühler, & Bickel, 2018).

However, most studies to date have concentrated on the concatenation of two call units (bigrams) (see e.g., Arnold & Zuberbühler, 2006; Coye, Ouattara, Zuberbühler, & Lemasson, 2015; Coye, Zuberbühler, & Lemasson, 2016; Collier et al., 2020) with such structures even being argued to represent the limit of animal combinatorial abilities (see e.g., Miyagawa & Clarke, 2019). Whether animal combinatorics lie beyond the two call level remains less explored, yet is critical for understanding how similar human and animal vocal systems are in their sequential dynamics.

We revisit this question through investigating the formation and internal structuring of call sequences in the highly social and cooperatively breeding common marmoset (*Callithrix Jacchus*). We recorded the vocal output of eight individuals (four pairs housed at the University of Zurich primate station, Switzerland, resulting in four female and four male subjects) via focal sampling.

This led to a corpus of 8600 calls, equalling 5800 call sequences (from unigram to 9-gram sequences), generated by the eight different call types present in the common marmoset vocal system. Through borrowing methods traditionally applied in computational linguistics, including simple frequency values, finite state automata (FSA), superordinate Markov models (MMs) (for an application in song systems see Honda and Okanoya, 1999; Hosino and Okanoya, 2000; Kershenbaum et al., 2014; Sainburg et al., 2019; Suzuki, Buck and Tyack, 2006; Kershenbaum et al., 2014) and collocation analysis, we show that common marmosets reliably concatenate vocalisations into larger structures of between two and nine calls (trigrams being the most common length of combinations).

Furthermore, using FSA and MMs, we identified sequential relationships between calls in combinations that point towards potential internal structuring. For example, trigram MMs highlighted a non-adjacent-like edge structure (a dependency between the first and last call element of a sequence (see Endress et al., 2009; Endress et al., 2010 and Grama, Wijnen & Kerkhoff, 2013)), whereby a trigrammic sequence starting with a food call always terminates with another food call in the last position. Multiple distinctive collocation analysis confirmed the food-call edge structure for three-call combinations and also four-call combinations. In addition, joint probability showed that the non-adjacent-like food-edge structure was present in 86% of all trigrams and 66% of all 4-grams.

While previous comparative research has identified non-adjacent structuring (structures that cannot be captured by more simple Markovian dynamics) in animal song systems (e.g., Sainburg et al. 2019; Suzuki, Buck and Tyack, 2006; Kershenbaum et al., 2014), this has not yet been shown in a non-human animal context-specific call system as is the case for the common marmosets' vocal communication system.

MMs and collocation analyses additionally provided some evidence for simple nestedness in common marmosets' call structures. For example, the bigram Tsk-Ek, which was identified by joint probability values and by the collocation analyses as one of the most common two-call units, also shows up in trigrammic combinations and when it does, it is reliably followed by a contact call giving rise to a "bigram-in-bigram" structuring.

We argue that call combinations in animals may be more sophisticated than previously thought and that the current dearth of evidence for larger call combinations in non-human animals could well be due to a lack of application of objective measurements to capture the sequential dynamics of structures at the call combination level.

References

Arnold, K., & Zuberbühler, K. (2006). Semantic combinations in primate calls. *Nature*, *441*(7091), 303–303.

Berthet, M., Mesbahi, G., Pajot, A., Cäsar, C., Neumann, C., & Zuberbühler, K. (2019). Titi monkeys combine alarm calls to create probabilistic meaning. Science Advances, 5(5), eaav3991.

Collier, K., Radford, A. N., Stoll, S., Watson, S. K., Manser, M. B., Bickel, B., & Townsend, S. W. (2020). Dwarf mongoose alarm calls: investigating a complex non-human animal call. Proceedings of the Royal Society B, 287(1935), 20192514.

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), 20150265.

Coye, C., Zuberbühler, K., & Lemasson, A. (2016). Morphologically structured vocalizations in female Diana monkeys. Animal Behaviour, 115, 97–105.

Endress A.D., Carden S., Versace E. & Hauser M.D. (2010). The apes' edge: positional learning in chimpanzees and humans. Animal cognition, 13(3), 483-495.

Grama I., Wijnen F. & Kerkhoff A. (2013). Constraints on non-adjacent dependency-learning: Distance matters an artificial grammar learning study with adults. In Boston University Conference on Language Development 37 online proceedings supplement, Boston, MA.

Honda, E., & Okanoya, K. (1999). Acoustical and syntactical comparisons between songs of the white-backed munia (Lonchura striata) and its domesticated strain, the Bengalese finch (Lonchura striata var. domestica). Zoological Science, 16(2), 319-326.

Hosino, T., & Okanoya, K. (2000). Lesion of a higher-order song nucleus disrupts phrase level complexity in Bengalese finches. Neuroreport, 11(10), 2091-2095.

Hurford, J. R. (2012). The origins of grammar: Language in the light of evolution II (Vol. 2). Oxford University Press.

Kershenbaum, A., Bowles, A. E., Freeberg, T. M., Jin, D. Z., Lameira, A. R., & Bohn, K. (2014). Animal vocal sequences: Not the Markov chains we thought they were. Proceedings of the Royal Society B: Biological Sciences, 281(1792), 20141370.

Leroux, M., & Townsend, S. W. (2020). Call combinations in great apes and the evolution of syntax. Animal Behavior and Cognition, 7(2), 131–139.

Miyagawa, S., & Clarke, E. (2019). Systems underlying human and old world monkey communication: One, two, or infinite. *Frontiers in psychology*, 1911.

Sainburg, T., Theilmann, B., Thielk, M., & Gentner, T. Q. (2019). Parallels in the sequential organization of birdsong and human speech. Nature Communications, 10(1), 1-11

Suzuki, R., Buck, J. R., & Tyack, P. L. (2006). Information entropy of humpback whale songs. The Journal of the Acoustical Society of America, 119(3), 1849–1866.

Townsend, S. W., Engesser, S., Stoll, S., Zuberbühler, K., & Bickel, B. (2018). Compositionality in animals and humans. PLoS Biology, 16(8), e2006425.