

Vocal learning, prosody, and basal ganglia: Don't underestimate their complexity¹

Andrea Ravignani,^a Mauricio Martins,^{a,b} and W. Tecumseh Fitch^a

^a*Department of Cognitive Biology, University of Vienna, A-1090 Vienna, Austria;*

^b*Language Research Laboratory, Lisbon Faculty of Medicine, 1649-028 Lisbon, Portugal.*

andrea.ravignani@univie.ac.at

mauricio.martins@univie.ac.at

tecumseh.fitch@univie.ac.at

<http://homepage.univie.ac.at/andrea.ravignani/>

www.researchgate.net/profile/Mauricio_Martins4/

<http://homepage.univie.ac.at/tecumseh.fitch/>

Abstract: Ackermann et al.'s arguments in the target article need sharpening and rethinking at both mechanistic and evolutionary levels. First, the authors' evolutionary arguments are inconsistent with recent evidence concerning nonhuman animal rhythmic abilities. Second, prosodic intonation conveys much more complex linguistic information than mere emotional expression. Finally, human adults' basal ganglia have a considerably wider role in speech modulation than Ackermann et al. surmise.

While Ackermann et al.'s theory is interesting, seems plausible, and may initially appear tempting, it is based on incomplete readings of several literatures. First, it is unclear why some of their arguments should only apply to the specific instances of rhythmic and prosodic control the authors discuss or why they *fail* to apply in other animal species. Their model assumes that

enhancement of in-group cooperation and cohesion was the main driving force for the evolution of speech via the intermediate step where vocal control and rhythm production would serve as chorusing and bonding tools. A key assumption is that speech would produce rhythmic abilities as an evolutionary by-product. This scenario is in line with some empirical observations (for reviews, see Fitch 2012; Geissmann 2000) and previous theoretical frameworks for the *origins of music* (Hagen & Bryant 2003; Hagen & Hammerstein 2009; Merker 2000; Merker et al. 2009). However, when applied to *language*, Ackermann et al.'s evolutionary model does not withstand cross-species validation: Many nonhuman animals exhibit rhythmic behaviors while lacking speech. Before primate rhythmic abilities can be compared with humans' at all, more evidence regarding flexibility in vocalizations' temporal patterning (Fedurek et al. 2013) and motor synchronization (Hattori et al. 2013) is needed in apes (cf. (Ravignani et al. 2013)).

Evidence from non-primate species also seems to undermine Ackermann et al.'s model. Two bird species, both vocal learners, have been shown to entrain to steady pulses (Hasegawa et al. 2011; Patel et al. 2009a), supporting Ackermann et al.'s model and Patel's hypothesis, whereby auditory-motor entrainment skills would be evolutionary by-products of vocal learning abilities (Patel 2006). However, recent evidence suggests that vocal learning and rhythmic abilities might be dissociated. Sea lions, unlike seals, show no evidence of vocal learning (Janik & Slater 1997) but nonetheless can reliably synchronize their movements to a range of musical stimuli at different tempi (Cook et al. 2013). Humans and sea lions are both rhythmically skilled, but only humans evolved vocal learning and speech. Therefore, sea lions constitute outliers inconsistent with the prediction of Ackermann et al.'s model. This species evolved cognitive rhythmic abilities, without evolving speech. Invoking additional evolutionary forces and

physiological mechanisms thus appears necessary: How can Ackermann et al.'s model be modified to avoid incorrectly predicting vocal learning in rhythmic-skilled species?

Second, Ackermann et al.'s model assumes that prosodic modulation of speech conveys mainly simple motivational-emotional information, and thus, that prosody and complex speech production had separate evolutionary histories. But evidence showing a tight connection between prosody and complex linguistic functions argues against this “double pathway” theory. Prosodic contour is influenced by syntactic constituent structure, semantic relations, phonological rhythm, pragmatic considerations, as well as by the length, complexity, and predictability of linguistic material (Wagner & Watson 2010). Furthermore, prosodic cues are used in childhood during acquisition of words (Christophe et al. 2008) and grammatical constructions (Männel et al. 2013), and in adulthood for syntactic processing (Christophe et al. 2008; Kjelgaard & Speer 1999; Langus et al. 2012; Wagner 2010) and word recognition (Cutler et al. 1997).

Contra Ackermann et al., such complex linguistic modulation of prosody seems to be a prerequisite for the acquisition and use of language, and this process is likely to be influenced by cognitive mechanisms specially modified in the human lineage. Comparative research on syntax precursors favors this hypothesis: The ability to assemble sequences of sounds into *hierarchical* patterns might be either human-specific, or very poorly developed in other species (Conway & Christiansen 2001; ten Cate & Okanoya 2012). Hence, developmental and comparative evidence point to a more complex cognitive integration of prosody and speech than allowed by the dual-pathway proposal of Ackermann et al. The challenge for Ackermann et al.'s theory is, therefore, to account for the modulation of prosody by human-specific cognitive functions (e.g., syntax),

which are clearly not evolutionary homologues of primate emotional vocalizations controlled by the anterior cingulate cortex.

Finally, Ackermann et al. propose an ontogenetic pathway in which: (1) basal ganglia (BG) are important to generate integrated templates of orofacial and laryngeal movements during childhood, but (2) in adulthood can be retrieved from cortical areas because these motor templates become well-trained. Later in ontogeny, BG would mostly subserve the modulation of emotional prosody, and not the coordination of speech production. These claims are not supported by currently available empirical data. For instance, Ackermann et al. cite Parkinson's Disease (PD) data to support their claims that, in adults, BG lesions only impair emotional prosody. In fact, PD patients with normal cognitive functioning are more impaired in semantic fluency tasks than in phonetic fluency (Henry & Crawford 2004). Additionally, contra Ackermann et al., BG subserve complex syntactic and semantic processing in adults, with empirical findings consistent across PD (Dominey & Inui 2009; Henry & Crawford 2004; Lewis et al. 1998), BG lesion (Kotz et al. 2003; Teichmann et al. 2008; Ullman et al. 1997), and neuroimaging research (Friederici & Kotz 2003). These data suggest that in adults the BG support multiple functions relevant to spoken language, not just simple emotional prosodic modulation.

Furthermore, contrary to the developmental pathway proposed by Ackermann et al., the acquisition of novel syntactic structures in adults depends on the medial temporal cortex, and the retrieval of syntactic templates *after* thorough learning mostly recruits the BG and perisylvian structures (Ullman 2004). This evidence shows that, contra Ackermann et al., BG are active in

the retrieval of over-learned procedures. Ackermann et al. therefore need to propose alternative explanations to reconcile child and adult data concerning the function of BG.

In conclusion, to make their model robust, Ackermann et al. must modify and refine their evolutionary and mechanistic explanations, and clarify which assumptions are necessary, and which are sufficient, for their explanatory framework to hold. Is their model robust enough to stand up to the clear, strong relationship between prosody and complex linguistic functions? How can Ackermann et al.'s model account for the complex functions of BG in adulthood? If in-group cohesion had to be achieved, why was precise vocal control specifically selected for, rather than general non-vocal rhythmic abilities? These and other questions need to be addressed if Ackermann et al.'s model is to become convincing.

1. Andrea Ravignani and Mauricio Martins contributed equally to this commentary as joint first authors.

ACKNOWLEDGMENTS

This work was supported by Fundação para a Ciência e Tecnologia grant SFRH/BD/64206/2009 (to Mauricio Martins) and European Research Council Advanced Grant 230604 SOMACCA (to Andrea Ravignani and W. Tecumseh Fitch).

References

- Christophe, A., Millotte, S., Bernal, S. & Lidz, J. (2008) Bootstrapping lexical and syntactic acquisition. *Language and Speech* 51(1–2):61–75. doi: 10.1177/00238309080510010501. [AR]
- Conway, C. M. & Christiansen, M. H. (2001) Sequential learning in non-human primates. *Trends in Cognitive Sciences* 5(12):539–46. [AR]
- Cook, P., Rouse, A., Wilson, M. & Reichmuth, C. (2013) A california sea lion (*Zalophus californianus*) can keep the beat: Motor entrainment to rhythmic auditory stimuli in a non-vocal mimic. *Journal of Comparative Psychology* 127(4):412–27. [AR]
- Cutler, A., Oahan, D. & van Donselaar, W. (1997) Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40(2):141–201. [AR]
- Dominey, P. F. & Inui, T. (2009) Cortico-striatal function in sentence comprehension: Insights from neurophysiology and modeling. *Cortex* 45(8):1012–18. doi: 10.1016/j.cortex.2009.03.007. [AR]
- Fedurek, P., Schel, A. M. & Slocombe, K. E. (2013) The acoustic structure of chimpanzee pant-hooting facilitates chorusing. *Behavioral Ecology and Sociobiology* 67(11):1781–89. [AR]
- Fitch, W. T. (2012) The biology and evolution of rhythm: Unraveling a paradox. In: *Language and music as cognitive systems*, ed. P. Rebuschat, M. Rohrmeier, J. A. Hawkins & I. Cross, pp. 73–95. Oxford University Press. [AR]
- Friederici, A. D. & Kotz, S. A. (2003) The brain basis of syntactic processes: Functional imaging and lesion studies. *NeuroImage* 20:S8–S17. doi: 10.1016/S1053-8119(03)00522-6. [AR]

- Geissmann, T. (2000) Gibbon songs and human music from an evolutionary perspective. In: *The origins of music*, ed. N. L. Wallin, B. Merker & S. Brown, pp. 103–23. MIT Press. [AR]
- Hagen, E. H. & Bryant, G. A. (2003) Music and dance as a coalition signaling system. *Human Nature* 14(1):21–51. [AR]
- Hagen, E. H. & Hammerstein, P. (2009) Did Neanderthals and other early humans sing? Seeking the biological roots of music in the territorial advertisements of primates, lions, hyenas, and wolves. *Musicae Scientiae* 13(Suppl. 2):291–320. [AR]
- Hasegawa, A., Okanoya, K., Hasegawa, T. & Seki, Y. (2011) Rhythmic synchronization tapping to an audio-visual metronome in budgerigars. *Scientific Reports* 1(Article 120):1–8. (Online journal). [AR]
- Hattori, Y., Tomonaga, M. & Matsuzawa, T. (2013) Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Scientific Reports* 3(Article 1566):1–6. (Online journal). [AR]
- Henry, J. D. & Crawford, J. R. (2004) Verbal fluency deficits in Parkinson’s disease: A meta-analysis. *Journal of the International Neuropsychological Society* 10:608–22. [AR]
- Janik, V. & Slater, P. J. B. (1997) Vocal learning in mammals. In: *Advances in the Study of Behavior*, vol. 26, ed. P. J. B. Slater, J. S. Rosenblatt, C. T. Snowdon, & M. Milinski, pp. 59–99. Academic Press. [AR]
- Kjelgaard, M. M. & Speer, S. R. (1999) Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language* 40:153–94. [AR]

- Kotz, S. A., Frisch, S., Cramon, S. Y. & Friederici, A. D. (2003) Syntactic language processing: ERP lesion data on the role of the basal ganglia. *Journal of the International Neuropsychological Society* 9:1053–60. [AR]
- Langus, A., Marchetto, E., Bion, R. A. H. & Nespors, M. (2012) Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language* 66(1):285–306. doi: 10.1016/j.jml.2011.09.004. [AR]
- Lewis, F. M., Lapointe, L. L., Murdoch, B. E. & Chenery, H. J. (1998) Language impairment in Parkinson's disease. *Aphasiology* 12(3):193–206. doi: 10.1080/02687039808249446. [AR]
- Männel, C., Schipke, C. S. & Friederici, A. D. (2013) The role of pause as a prosodic boundary marker: Language ERP studies in German 3- and 6-year-olds. *Developmental Cognitive Neuroscience* 5:86–94. doi: 10.1016/j.dcn.2013.01.003. [AR]
- Merker, B. (2000) Synchronous chorusing and the origins of music. *Musicae Scientiae* 3(Suppl. 1):59–73. [AR]
- Merker, B., Madison, G. & Eckerdal, P. (2009) On the role and origin of isochrony in human rhythmic entrainment. *Cortex* 45(1):4–17. [AR]
- Patel, A. D. (2006) Musical rhythm, linguistic rhythm, and human evolution. *Music Perception: An Interdisciplinary Journal* 24(1):99–104. [AR]
- Patel, A. D., Iversen, J. R., Bregman, M. R. & Schulz, I. (2009a) Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology* 19(10):827–30. [AR]
- Ravignani, A., Olivera, V. M., Gingras, B., Hofer, R., Hernández, C. R., Sonnweber, R.-S. & Fitch, W. T. (2013) Primate drum kit: A system for studying acoustic pattern production

- by non-human primates using acceleration and strain sensors. *Sensors* 13(8):9790–820. [AR]
- Teichmann, M., Gaura, V., Demonet, J. F., Supiot, F., Delliaux, M., Verny, C., Renou, P., Remy, P. & Bachoud-Levi, A. C. (2008) Language processing within the striatum: Evidence from a PET correlation study in Huntington’s disease. *Brain* 131(4):1046–56. doi: 10.1093/brain/awn036. [AR]
- ten Cate, C. & Okanoya, K. (2012) Revisiting the syntactic abilities of non-human animals: Natural vocalizations and artificial grammar learning. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367(1598):1984–94. [AR]
- Ullman, M. T. (2004) Contributions of memory circuits to language: The declarative/procedural model. *Cognition* 92(1–2):231–70. [AR]
- Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J. & Pinker, S. (1997) A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience* 9(2):266–76. [AR]
- Wagner, M. (2010) Prosody and recursion in coordinate structures and beyond. *Natural Language & Linguistic Theory* 28(1):183–237. doi: 10.1007/s11049-009-9086-0. [AR]
- Wagner, M. & Watson, D. G. (2010) Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes* 25(7–9):905–45. doi: 10.1080/01690961003589492. [AR]