

COMPARATIVE COGNITION & BEHAVIOR REVIEWS

Toward Interdisciplinary Integration in the Study of Comparative Cognition: Insights from Studying the Evolution of Multimodal Communication

Elizabeth Qing Zhang

Jiangsu Normal University

Michael Pleyer

Nicolaus Copernicus University in Toruń

In this article, we highlight the importance of interdisciplinary integration in the study of comparative cognition. Specifically, we argue that the study of comparative cognition can benefit from broadening its focus and integrating information from diverse subfields and including collaborations from other fields. We take the evolution of multimodal communication as an example to illustrate that an interdisciplinary integration of linguistics, animal behavior, cognitive neuroscience, and genetics provides a more comprehensive picture of this phenomenon.

Keywords: multimodal communication, interdisciplinary integration, linguistics, animal behavior, cognitive neuroscience

One of the most promising advancements in comparative cognition in recent years is that comparative psychologists have shifted to working on a broader range of species and cognitive abilities and have emphasized a team-science approach (e.g., Guillette & Sturdy, 2020). However, a remaining challenge for comparative cognition as a whole is that it has not paid enough attention to the integration of information from different disciplines. Recently, there have increasingly been calls for an “integrative comparative cognition,” such as by Burmeister and Liu (2020). They argued that neurobiological and neurogenomic studies can shed important light on the cognitive phenotypes that are the subject of comparative cognition. In this article, we highlight the importance of taking a broader range of fields into consideration using the evolution of multimodal communication as an example

to illustrate the benefits this interdisciplinary approach would have on the field of comparative cognition.

Human communication is fundamentally multimodal. Vocal and visual cues are integrated in human linguistic interaction—for example, in phenomena such as co-speech gestures and facial expressions that accompany vocalizations in spoken language, as well as the iconicity, sound symbolism, and cross-modal correspondences that motivate many aspects of language structure in both spoken and signed languages (e.g., Dingemanse et al., 2015; Vigliocco et al., 2014). Appreciation of human communication as a multimodal phenomenon supports the idea that language itself has a multimodal origin (e.g., Fröhlich et al., 2019; Levinson & Holler, 2014). Indeed, comparative research on nonhuman animals shows that multimodality is a ubiquitous property of many animal communication

systems (Ota et al., 2015; Partan & Marler, 1999), suggesting evolutionary continuity and deep evolutionary roots of multimodal communication.

However, studies on the evolution of multimodal communication have focused mainly on studying nonhuman primates (e.g., Fröhlich & van Schaik, 2018; Liebal et al., 2014), likely because of their close evolutionary relationship to humans. Nonhuman primates show simultaneous production of communication signals in the manual, facial, and vocal modalities (Genty et al., 2014; Micheletta et al., 2013). A study of chimpanzees found that about half of all vocalizations were produced in combination with another communicative modality (Tagliatalata et al., 2015). On the other hand, existing studies on multimodal communication in diverse species suggest that multimodality has even deeper evolutionary roots. However, although a wealth of studies are on multimodal communication in different species, few studies have explicitly addressed the evolutionary continuity of multimodality outside of primates. That is, so far, studies showing the ubiquity of multimodality in the animal kingdom have not been properly integrated into an account of the evolution of multimodal communication. Furthermore, apart from the behavioral and cognitive levels, neuroscientific and genetic (genomic) studies can also provide revealing insights into the evolutionary continuity of multimodal communication.

For this reason, we argue that the study of the evolution of multimodal communication is in need of interdisciplinary integration, which we believe is an important future challenge for the field of comparative cognition and behavior.

On one hand, the field requires a combination of various research fields that explore the role of multimodality in humans in both naturalistic and laboratory settings (e.g., Macuch Silva et al., 2020; Rasenberg et al., 2022), as well as research in comparative cognition on the role of multimodality in a wide range of nonhuman species. For

example, combinations of tactile, olfactory, acoustic, and visual cues have been reported in fruit flies (Ewing, 1983), fish (Tavolga, 1956), and birds (Dalziell et al., 2013; Ota et al., 2015). In songbirds, courtship displays integrate songs with hops, head motions and beak movements (Williams, 2001). In addition, song type repertoire is coordinated temporally with a dance-like movement repertoire (Dalziell & Peters, 2013). Overall, these data indicate that multimodal communication has a deep phylogenetic origin dating back to invertebrates.

On the other hand, the study of the evolution of multimodality would also profit from interdisciplinary insight from neuroscience and genetics. From a neuroscientific perspective, the hippocampus and basal ganglia both represent conserved subcortical structures that are found in all vertebrates. Homologous neural structures have also been proposed for invertebrates (Lin et al., 2013; Wolff & Strausfeld, 2015). The basal ganglia are mostly involved in action selection, motor control, and cognitive functions such as procedural learning and memory (Graybiel, 2005). Functions of the hippocampus include declarative learning and memory, navigation, and episodic memory (Voss et al., 2017). Concerning communication, studies on vocal production learning in animals, especially songbirds, have demonstrated a crucial role for the basal ganglia in song learning (Jarvis, 2019). Detailed comparisons of the neural circuitry of songbirds and humans have also shown that certain analogous (potentially homologous) cortico-basal ganglia-thalamo-cortical circuits are essential to vocal learning (Pfenning et al., 2014). Regarding the hippocampus, studies on human patients with amnesia suggest that it is also vital for the production of gestures (Hilliard et al. 2017). As the hippocampus is involved in spatial cognition, this suggests a hippocampal contribution to the evolution of the use of gestures in humans (Levinson, 2023). Therefore, there is suggestive evidence that the connection between the hippocampus and the basal ganglia could underlie multimodal communication across species. Still, more interdisciplinary work is needed in this domain, representing an important challenge for future work on the evolution of multimodal communication.

Last, genetic studies have the potential to serve as an important puzzle piece in unraveling the evolution of multimodal communication from an interdisciplinary perspective. As a case in point, the integration of speech and gesture might be influenced by the human version of the FOXP2¹

1. We follow the convention that capitalized FOXP2 “refers to the human gene, Foxp2 refers to the gene in mice and FoxP2 refers to the gene in other species” (Schattton & Scharff, 2017, p. 26).

Author Note: Elizabeth Qing Zhang, School of Linguistic Sciences and Arts, Jiangsu Normal University, Shanghai Road 101, Tongshan New District, 221116, Xuzhou, Jiangsu, China.

Correspondence concerning this article should be addressed to Elizabeth Qing Zhang at zqelizab@163.com.

Acknowledgments: Michael Pleyer was supported by project No. 2021/43/P/HS2/02729 co-funded by the National Science Centre and the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 945339.

gene. In evolution, FoxP2 represents a conserved transcription factor among vertebrates; there is also indicative data in invertebrates. *Drosophila* possesses a homolog of FoxP2: FoxP, which is responsible for sex-specific walking and flight as well as pulse-song structure (Lawton et al., 2014). Studies in vertebrates also indicate a connection of FoxP2 to both the basal ganglia and hippocampus, and their interaction in multimodal communication. When Foxp2 is knocked out in mice, infant mice will produce abnormal ultrasonic vocalizations (Shu et al., 2005). Further studies in mice have demonstrated that heterozygous mutations of Foxp2 impair sensorimotor association learning (Kurt et al., 2012). Also in mice, Zbtb20, a repressor of Foxp2, has been found to bind to and repress cortical layer marker genes (including Foxp2) in the developing hippocampus (Nielsen et al., 2014). Knockdown studies in songbirds, in which the expression of particular genes is reduced, also show a connection of FoxP2 to vocalizations. Specifically, if FoxP2 expression is knocked down in Area X in juvenile zebra finches, this affects the completeness and accuracy of song production learning (Haesler et al., 2007). More research is needed to untangle the possible influence of human FOXP2 on the evolution of specifically human multimodal communication. Human FOXP2 has incorporated two fixed amino acid changes in a broadly defined transcription suppression domain (Zhang et al., 2002). These two amino acid changes (N325S, T303N) occurred at some point since the evolutionary split from the lineage of chimpanzees and bonobos (Enard et al., 2002) and were likely present before the split of Neanderthals and *Homo sapiens* (Krause et al., 2007), which is currently estimated to have happened between 800 thousand years ago and 400 thousand years ago (cf. Endicott et al., 2010; Harvati & Reyes-Centeno, 2022). This suggests evolutionary continuity of (humanlike) multimodal communication, possibly dating back to *Homo heidelbergensis* (Dediu & Levinson, 2018). However, an ongoing debate concerns the timing of the evolution of FOXP2 and other possible subtle changes that might have occurred since the split from the Neanderthal lineage (see, e.g., Fisher, 2019). Animal studies offer important further insights here, as mice injected with a humanized version of FOXP2 showed a reduced dopamine level, increased dendritic length, and long-term synaptic depression (Enard et al., 2009), suggesting a role of human FOXP2 in altering the basal ganglia structure and function. Moreover, mice with a humanized version of FOXP2 also show an accelerated transition from declarative learning to procedural learning (Schreiweis et al., 2014). As the neural bases of declarative and procedural performance are the hippocampus and basal ganglia, respectively, this suggests

a key role of FOXP2 in better connecting the basal ganglia and hippocampus; this connection represents an important aspect of human multimodal communication.

The main thrust of this commentary rests on two aspects: On one hand, it represents a call to take multimodality seriously in the study of communication in different species and to include a wider range of species in comparative cognition studies of multimodal communication. On the other hand, it is a call for interdisciplinary integration. Using the evolution of multimodal communication as an example, we want to make a case for the idea that comparative cognition can benefit from broadening its focus, integrating information from different subfields and including collaborators from other fields. As we have shown, the integration of insights from fields such as the language sciences, animal communication, neuroscience, and genetics has the potential to make important contributions to the study of multimodality and its evolution. We hope that, in the future, such interdisciplinary integration will lead to further exciting discoveries and the development of interspecies frameworks for the study of multimodal communication. More generally, our discussion of the evolution of multimodal communication serves as an example of how broader interdisciplinary collaboration within and outside comparative cognition can potentially greatly move the field forward.

References

- Burmeister, S. S., & Liu, Y. (2020). Integrative comparative cognition: Can neurobiology and neurogenomics inform comparative analyses of cognitive phenotype? *Integrative and Comparative Biology*, *60*(4), 925–928. <https://doi.org/10.1093/icb/icaa113>
- Dalziell, A. H., Peters, R. A., Cockburn, A., Dorland, A. D., Maisey, A. C., & Magrath, R. D. (2013). Dance choreography is coordinated with song repertoire in a complex avian display. *Current Biology*, *23*(12), 1132–1135. <https://doi.org/10.1016/j.cub.2013.05.018>
- Dediu, D., & Levinson, S. C. (2018). Neanderthal language revisited: Not only us. *Current Opinion in Behavioral Sciences*, *21*, 49–55. <https://doi.org/10.1016/j.cobeha.2018.01.001>
- Dingemans, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, *19*(10), 603–615. <https://doi.org/10.1016/j.tics.2015.07.013>

- Enard, W., Gehre, S., Hammerschmidt, K., Hölter, S. M., Blass, T., Somel, M., ... Pääbo, S. (2009). A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell*, *137*(5), 961–971. <https://doi.org/10.1016/j.cell.2009.03.041>
- Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., ... Pääbo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, *418*(6900), 869–872. <https://doi.org/10.1038/nature01025>
- Endicott, P., Ho, S. Y., & Stringer, C. (2010). Using genetic evidence to evaluate four palaeoanthropological hypotheses for the timing of Neanderthal and modern human origins. *Journal of Human Evolution*, *59*(1), 87–95. <https://doi.org/10.1016/j.jhevol.2010.04.005>
- Ewing, A. W. (1983). Functional aspects of drosophila courtship. *Biological Reviews*, *58*(2), 275–292. <https://doi.org/10.1111/j.1469-185X.1983.tb00390.x>
- Fisher, S. E. (2019). Human genetics: The evolving story of FOXP2. *Current Biology*, *29*(2), R65–R67. <https://doi.org/10.1016/j.cub.2018.11.047>
- Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & van Schaik, C. P. (2019). Multimodal communication and language origins: Integrating gestures and vocalizations. *Biological Reviews*, *94*(5), 1809–1829. <https://doi.org/10.1111/brv.12535>
- Fröhlich, M., & van Schaik, C. P. (2018). The function of primate multimodal communication. *Animal Cognition*, *21*, 619–629. <https://doi.org/10.1007/s10071-018-1197-8>
- Genty, E., Clay, Z., Hobaiter, C., & Zuberbühler, K. (2014). Multi-modal use of a socially directed call in bonobos. *PLoS ONE*, *9*(1), Article e84738. <https://doi.org/10.1371/journal.pone.0084738>
- Graybiel, A. M. (2005). The basal ganglia: Learning new tricks and loving it. *Current Opinion in Neurobiology*, *15*(6), 638–644. <https://doi.org/10.1016/j.conb.2005.10.006>
- Guillette, L. M., & Sturdy, C. B. (2020). Unifying psychological and biological approaches to understanding animal cognition. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *74*(3). <https://doi.org/10.1037/cep0000233>
- Haesler, S., Rochefort, C., Georgi, B., Licznarski, P., Osten, P., & Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus area X. *PLoS Biology*, *5*(12), 2885–2897. <https://doi.org/10.1371/journal.pbio.0050321>
- Harvati, K., & Reyes-Centeno, H. (2022). Evolution of homo in the middle and late Pleistocene. *Journal of Human Evolution*, *173*, Article 103279. <https://doi.org/10.1016/j.jhevol.2022.103279>
- Hilliard, C., Cook, S. W., & Duff, M. C. (2016). Hippocampal declarative memory supports gesture production: Evidence from amnesia. *Cortex*, *85*, 25–36. <https://doi.org/10.1016/j.cortex.2016.09.015>
- Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. *Science*, *366*(6461), 50–54. <https://doi.org/10.1126/science.aax0287>
- Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R. E., Burbano, H. A., ... Pääbo, S. (2007). The derived FOXP2 variant of modern humans was shared with Neandertals. *Current Biology*, *17*(21), 1908–1912. <https://doi.org/10.1016/j.cub.2007.10.008>
- Kurt, S., Fisher, S. E., & Ehret, G. (2012). Foxp2 mutations impair auditory-motor association learning. *PLoS ONE*, *7*(3), Article e33130. <https://doi.org/10.1371/journal.pone.0033130>
- Lawton, K. J., Wassmer, T. L., & Deitcher, D. L. (2014). Conserved role of Drosophila melanogaster FoxP in motor coordination and courtship song. *Behavioural Brain Research*, *268*, 213–221. <https://doi.org/10.1016/j.bbr.2014.04.009>
- Levinson, S. C. (2023). Gesture, spatial cognition and the evolution of language. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *378*(1875), Article 20210481. <https://doi.org/10.1098/rstb.2021.0481>
- Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), Article 20130302. <https://doi.org/10.1098/rstb.2013.0302>

- Liebal, K., Waller, B. M., Burrows, A. M., & Slocombe, K. E. (2014). *Primate communication: A multimodal approach*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139018111>
- Lin, C. Y., Chuang, C. C., Hua, T. E., Chen, C. C., Dickson, B. J., Greenspan, R. J., & Chiang, A. S. (2013). A comprehensive wiring diagram of the protocerebral bridge for visual information processing in the *Drosophila* brain. *Cell Reports*, 3(5), 1739–1753. <https://doi.org/10.1016/j.celrep.2013.04.022>
- Macuch Silva, V., Holler, J., Ozyurek, A., & Roberts, S. G. (2020). Multimodality and the origin of a novel communication system in face-to-face interaction. *Royal Society Open Science*, 7(1), Article 182056. <https://doi.org/10.1098/rsos.182056>
- Micheletta, J., Engelhardt, A., Matthews, L., Agil, M., & Waller, B. M. (2013). Multicomponent and multimodal lip smacking in crested macaques (*Macaca nigra*). *American Journal of Primatology*, 75(7), 763–773. <https://doi.org/10.1002/ajp.22105>
- Nielsen, J. V., Thomassen, M., Møllgård, K., Noraberg, J., & Jensen, N. A. (2014). Zbtb20 defines a hippocampal neuronal identity through direct repression of genes that control projection neuron development in the isocortex. *Cerebral Cortex*, 24(5), 1216–1229. <https://doi.org/10.1093/cercor/bhs400>
- Ota, N., Gahr, M., & Soma, M. (2015). Tap dancing birds: The multimodal mutual courtship display of males and females in a socially monogamous songbird. *Scientific Reports*, 5(1), Article 16614. <https://doi.org/10.1038/srep16614>
- Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science*, 283(5406), 1272–1273. <https://doi.org/10.1126/science.283.5406.1272>
- Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., Howard, J. T., Wirthlin, M., Lovell, P. V., Ganapathy, G., Mouncastle, J., Moseley, M. A., Thompson, J. W., Soderblom, E. J., Iriki, A., Kato, M., Gilbert, M. T. P., Zhang, G., Bakken, T., ... Jarvis, E. D. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science*, 346(6215), Article 1256846. <https://doi.org/10.1126/science.1256846>
- Rasenberg, M., Pouw, W., Özyürek, A., & Dingemanse, M. (2022). The multimodal nature of communicative efficiency in social interaction. *Scientific Reports*, 12(1), Article 19111. <https://doi.org/10.1038/s41598-022-22883-w>
- Schatton, A., & Scharff, C. (2017). Next stop: Language. The ‘FOXP2’ gene’s journey through time. *Métode Science Studies Journal*, 7, 25–33. <https://doi.org/10.7203/metode.7.7248>
- Schreiweis, C., Bornschein, U., Burguière, E., Kerimoglu, C., Schreiter, S., Dannemann, M., ... Graybiel, A. M. (2014). Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proceedings of the National Academy of Sciences*, 111(39), 14253–14258. <https://doi.org/10.1073/pnas.1414542111>
- Shu, W., Cho, J. Y., Jiang, Y., Zhang, M., Weisz, D., Elder, G. A., ... Buxbaum, J. D. (2005). Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27), 9643–9648. <https://doi.org/10.1073/pnas.0503739102>
- Tagliatela, J. P., Russell, J. L., Pope, S. M., Morton, T., Bogart, S., Reamer, L. A., ... Hopkins, W. D. (2015). Multimodal communication in chimpanzees. *American Journal of Primatology*, 77(11), 1143–1148. <https://doi.org/10.1002/ajp.22449>
- Tavolga, W. N. (1956). Visual, chemical and sound stimuli as cues in the sex discriminatory behavior of the gobiid fish *Bathygobius soporator*. *Zoologica*, 41(2), 49–64. <https://doi.org/10.5962/p.203402>
- Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: Implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), Article 20130292. <https://doi.org/10.1098/rstb.2013.0292>
- Voss, J. L., Bridge, D. J., Cohen, N. J., & Walker, J. A. (2017). A closer look at the hippocampus and memory. *Trends in Cognitive Sciences*, 21(8), 577–588. <https://doi.org/10.1016/j.tics.2017.05.008>
- Williams, H. (2001). Choreography of song, dance and beak movements in the zebra finch (*Taeniopygia guttata*). *The Journal of Experimental Biology*, 204(Pt. 20), 3497–506. <https://doi.org/10.1242/jeb.204.20.3497>

- Wolff, G. H., & Strausfeld, N. J. (2015). Genealogical correspondence of mushroom bodies across invertebrate phyla. *Current Biology*, *25*(1), 38–44. <https://doi.org/10.1016/j.cub.2014.10.049>
- Zhang, J., Webb, D. M., & Podlaha, O. (2002). Accelerated protein evolution and origins of human-specific features: FOXP2 as an example. *Genetics*, *162*(4), 1825–1835. <https://doi.org/10.1093/genetics/162.4.1825>