DOI:10.3819/CCBR.2024.190001 Volume 19, 2024

Comparative Cognition & Behavior Reviews

Comparative Cognition Needs Big Team Science: How Large-Scale Collaborations Will Unlock the Future of the Field

Nicolás Alessandroni[1,*], Drew Altschul[2,3], Marina Bazhydai[4], Krista Byers-Heinlein[1], Mahmoud Elsherif[5,6], Biljana Gjoneska[7], Ludwig Huber[8], Valeria Mazza[9,10], Rachael Miller[11,12], Christian Nawroth[13], Ekaterina Pronizius[14], Muhammad A. J. Qadri[15], Vedrana Šlipogor[16], Melanie Soderstrom[17], Jeffrey R. Stevens[18], Ingmar Visser[19], Madison Williams[1], Martin Zettersten[20], & Laurent Prétôt[21]

[1]Concordia University, Canada, [2]University of Edinburgh, Scotland, [3]Scottish Primate Research Group, Scotland, [4]Lancaster University, England, [5]University of Leicester, England, [6]University of Birmingham, England, [7]Macedonian Academy of Sciences and Arts, North Macedonia, [8]University of Veterinary Medicine Vienna, Austria, [9]University of Tuscia, Italy, [10]University of Potsdam, Germany, [11]Anglia Ruskin University, England, [12]University of Cambridge, England, [13]Research Institute for Farm Animal Biology, Germany, [14]University of Vienna, Austria, [15]College of the Holy Cross, USA, [16]University of South Bohemia, Czech Republic, [17]University of Manitoba, Canada, [18]University of Nebraska-Lincoln, USA, [19]University of Amsterdam, The Netherlands, [20]Princeton University, USA, [21]Pittsburg State University, USA

Comparative cognition research has been largely constrained to isolated facilities, small teams, and a limited number of species. This has led to challenges such as conflicting conceptual definitions and underpowered designs. Here, we explore how Big Team Science (BTS) may remedy these issues. Specifically, we identify and describe four key BTS advantages — increasing sample size and diversity, enhancing task design, advancing theories, and improving welfare and conservation efforts. We conclude that BTS represents a transformative shift capable of advancing research in the field.

Keywords: Big Team Science, comparative cognition, large-scale collaboration, metascience, research culture

Comparative cognition research aims to describe and explain the evolutionary background and functions of cognitive skills that allow individuals across species to adapt and act in their environment (Tinbergen, 1963). Despite advancements in the discipline, several long-standing questions remain unanswered—for example, how cognitive variability across species relates to environmental, social, and genetic factors. Elements hampering progress in comparative cognition parallel those encountered in other disciplines and often relate to the ongoing *reproducibility crisis*, including (a) variations in the definition of concepts across fields, (b) small sample sizes and underpowered designs, (c) constraints on research resources, (d) fragmented research efforts, (e) a lack of protocols for handling suboptimal data, and (f) (sometimes unknown) differences in laboratory practices (see Vazire, 2018).

Recently, Big Team Science (BTS) has become popularized as a way of addressing these limitations. BTS is a grassroots approach to research, when large numbers of researchers join forces and pool their resources and

efforts to answer crucial questions in their field, either in a single species (e.g., *ManyBabies, ManyDogs*) or across multiple species (e.g., *ManyPrimates*, *ManyBirds, ManyManys*). Scholars have stipulated numerous benefits of BTS (e.g., Forscher et al., 2023). Most notably, BTS enables researchers to amass substantially larger samples as a collective group than any one research team could gather independently. For example, in the field of infant studies, BTS has increased sample sizes by two orders of magnitude, from dozens to thousands of participants in a single study (see ManyBabies Consortium, 2020). Larger samples increase the statistical power of analyses, improve external validity, and yield more precise effect size estimates. At the same time, large high-quality datasets provide opportunities for studying secondary research questions. BTS may also make the research process more inclusive for underresourced, underrepresented, and early-career researchers with limited access to subjects and equipment. Furthermore, BTS allows for enhanced communication and networking among researchers working across countries and institutions (e.g., ManyDogs Project, 2023). By breaking down research silos, BTS provides infrastructure (i.e., a distributed network of sites) for streamlining methodologies (e.g., design, data analysis) and fostering best research practices, thus addressing replication concerns and spurring incremental and systematic scientific improvement.

Author Note: *Nicolás Alessandroni, Concordia Infant Research Lab, Department of Psychology, Concordia University (Canada), 7141 Sherbrooke St. West, PY-033, Montréal, Québec, H4B 1R6, Canada.*

Correspondence concerning this article should be addressed to Nicolás Alessandroni at nicolas.alessandroni@concordia.ca.

Acknowledgments: *The authors of this work are listed alphabetically with the exception of Nicolás Alessandroni (first and corresponding author) and Laurent Prétôt (last author), who coordinated and managed the preparation of the manuscript in their capacity as co-leads of the ManyManys 1 Big Team Science collaboration. The authors have no known conflict of interest to disclose.*

Code Availability Statement: *The code used to generate Figure 1 is openly available at https://osf.io/mavku/.*

Credit Report: *Conceptualization: Nicolás Alessandroni and Laurent Prétôt. Project administration: Nicolás Alessandroni and Laurent Prétôt. Supervision: Nicolás Alessandroni and Laurent Prétôt. Visualization: Nicolás Alessandroni. Writing—original draft: Nicolás Alessandroni, Drew Altschul, Marina Bazhydai, Krista Byers-Heinlein, Mahmoud Elsherif, Biljana Gjoneska, Ludwig Huber, Valeria Mazza, Rachael Miller, Christian Nawroth, Ekaterina Pronizius, Muhammad A. J. Qadri, Vedrana Šlipogor, Melanie Soderstrom, Jeffrey R. Stevens, Ingmar Visser, Madison Williams, Martin Zettersten, and Laurent Prétôt. Writing—Review and editing: Nicolás Alessandroni, Drew Altschul, Marina Bazhydai, Krista Byers-Heinlein, Mahmoud Elsherif, Biljana Gjoneska, Ludwig Huber, Valeria Mazza, Rachael Miller, Christian Nawroth, Ekaterina Pronizius, Muhammad A. J. Qadri, Vedrana Šlipogor, Melanie Soderstrom, Jeffrey R. Stevens, Ingmar Visser, Madison Williams, Martin Zettersten, and Laurent Prétôt.*

Funding: *This work was supported by the following fellowships and grants: FRQSC Postdoctoral Fellowship [NA], British Academy PF20/100086 [DA], Leverhulme Early-Career Research Fellowship [ME], Austrian Science Funds W1262 [LH], University of South Bohemia Postdoctoral Fellowship [VŠ], SSHRC Partnership Development Grant 890-2020-0059 [MS], NIH NICHD F32HD110174 [MZ], National Institute of General Medical Sciences of the National Institutes of Health P20 GM103418 [LP].*

What Can Comparative Cognition Specifically Gain from Embracing BTS?

In this section, we explore key benefits that comparative cognition research can gain from BTS. The authors, representing *ManyManys*—a recently formed BTS collaboration on comparative cognition and behavior across animal taxa (https://manymanys.github.io)—draw upon our collective experience and discussions to shed light on the transformative potential of this collaborative methodology.

Enhanced Sample Size and Diversity

One fundamental goal of comparative research is to compare results across a variety of subjects and species. Yet individual labs are often limited to a few sites, apparatuses, species, strains or breeds, and even subjects. These limitations diminish the generalizability of findings, hindering progress. By sharing the costs of data collection—in terms of resources, time, and expertise—across groups of researchers spanning different countries and settings, BTS naturally fosters increased sample sizes and diversity (e.g., by allowing wider access to underrepresented or endangered species), thereby catalyzing the assembly of high-value datasets (Figure 1). This cost-sharing is key for generating the statistical power needed to assess the small effects often found in comparative research, thus reducing false positives and negatives and enabling more precise depictions of phenomena (Farrar et al., 2020).

At the same time, comparative data are uniquely complex because of their hierarchical nature. Indeed, in any multitaxa study (e.g., MacLean et al., 2014), measuring *within-species*, *between-species*, and *between-taxa* stochastic variation is crucial to assess their possible influence on the analyses of interest (e.g., within-species individual differences or taxon-specific variations that, although not the primary focus, can affect research outcomes). In statistical models, these sources of variability within and between groups can be addressed only with a sufficiently large sample size at each level. Without such information, the field is left with an incomplete understanding of the true patterns and principles underlying the phenomena under study.

Sample diversity is a related challenge in comparative cognition, as the majority of studies have historically concentrated on a relatively limited set of species accessible to researchers. To illustrate, recent reviews of studies on primate and avian cognition and behavior reveal that only 68 of more than 500 primate species and 141 of more than 10,000 avian species have been examined (Lambert et al., 2022; ManyPrimates et al., 2022). These studies have featured median within-species sample sizes of 7 for primates and 14 for birds. In addition, between-species comparisons have been infrequent, with just 19% of primate studies and 10.9% of avian studies testing more than one species. Similar limitations exist for other taxa where available data are so scarce that comprehensive reviews are uninformative.

In addition to increasing the range of species under investigation, researchers can enrich sample diversity by considering other factors, including subtle differences in the research process stemming from the accumulation of (nonreported) specific lab setups, practices, and idiosyncrasies and the variety of keeping conditions and rearing histories. Notably, comparative cognition research

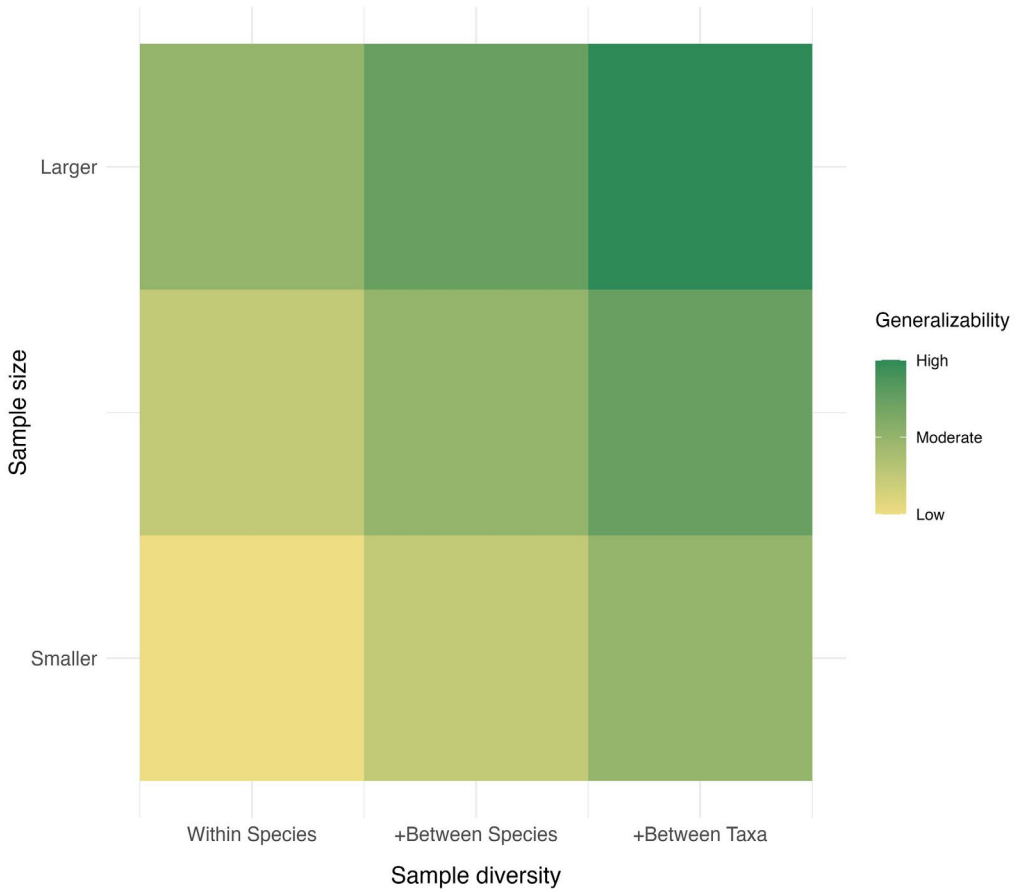

Figure 1. A heatmap illustrating, in principle, how the generalizability of findings relates to sample size and sample diversity. The horizontal (*x*) axis represents different ranges of sample diversity by considering different sources of variability (within-species, between-species, and between-taxa). The vertical (*y*) axis represents different ranges of sample size. The color shows the overall theoretical generalizability of findings if we assume that sample size and sample diversity have the same impact on generalizability.

has tended to favor samples originating from so-called BIZARRE settings (Barren, Institutional, Zoo, And other Rare Rearing Environments; Leavens et al., 2010) and has been susceptible to STRANGE-related biases (Social background, Trappability and self-selection, Rearing history, Acclimation and habituation, Natural changes in responsiveness, Genetic makeup, and Experience; Webster & Rutz, 2020). A promising strategy to address these limitations is to expand the number of research facilities housing various populations of the same species and provide access to a broader array of species, thereby enriching sample diversity and limiting the degree to which idiosyncrasies of individual sites can bias results (see Voelkl et al., 2018).

Balancing Cross-Species Standardization and Species-Fair Design

Comparing animal cognition across taxa is a challenging endeavor. First, individuals from different species vary in terms of size, anatomy, physiology, and skills, among other dimensions. Second, any given experiment can take on many different forms depending on aspects such as task modality (e.g., manual vs. computerized), testing environment (e.g., presence vs. absence of conspecifics), population (e.g., captive vs. wild), cue type (e.g., size, color), and task length. As a result, certain methodological approaches are more feasible with some taxa than with others. BTS projects can test a larger suite of species and more diverse sets of subpopulations than traditional projects can accommodate, which results in increased relevance and complexity of methodological considerations. Arguably, BTS makes existing challenges within the comparative literature more obvious.

These challenges raise practical questions of how to best measure between-taxa variability in a way that makes comparisons interpretable: Which procedural parameters should remain consistent across species, and which should be allowed to vary? One approach is to maximize cross-species standardization by testing all subjects on as similar a paradigm as possible, which, as a principle, should enhance the internal validity of studies. However, the one-size-fits-all approach (i.e., requiring individuals from different species to perform the same task) can lead to variable overcontrolling and potential experimental failure rather than improving internal validity in any significant way. Moreover, keeping experimental conditions constant across species can range from being daunting to virtually impossible. A second approach is that of *species-fair comparisons* (see Farrar et al., 2021). This approach is anchored in a persuasive tenet: allowing diverse organisms to demonstrate their knowledge calls for accommodating their performance constraints, thereby ensuring that their outcomes remain impartial to these constraints (Firestone, 2020).

Ultimately, BTS in comparative cognition will need to strike the right balance between cross-species standardization and species-fair design. Fortunately, BTS collaborations bring together an atypically large number of scientists—usually with heterogeneous backgrounds—which turns them into unparalleled think tanks distinctly equipped to address methodological concerns. By judiciously standardizing what is reasonable to keep constant while allowing certain parameters to vary in how a task is implemented across species (e.g., based on species-specific needs and preferences), researchers are better equipped to achieve a delicate balance between standardization and tailoring, ultimately promoting a more robust and nuanced understanding of cognitive abilities in diverse organisms (Tecwyn, 2021). Striking this balance may be one of the most difficult problems in comparative research. BTS brings together the research expertise needed to help address this problem while providing the infrastructure to directly test the impact of species-fair versus strictly standardized designs.

Theoretical Advancements

In comparative cognition, much like in other scientific disciplines, researchers grapple with competing theories seeking to explain the same cognitive phenomena. One obstacle hindering the endorsement of any single theory can be the scarcity of empirically robust evidence. For example, several ongoing debates revolve around whether nonhuman species—mainly invertebrates—genuinely experience *full-fledged emotions* or merely exhibit *emotion-like behaviors* (e.g., Solvi et al., 2016). However, disagreement can also arise from radically opposed conceptual and epistemological worldviews and research traditions (Bitterman, 1975; Castorina, 2021). This is particularly relevant to comparative cognition and its unique commitment to exploring whether overarching principles governing cognition and behavior exist across taxa. To further compound the issue, comparative cognition is also inherently multidisciplinary, given the breadth of methodologies and research topics relevant across taxa.

BTS amplifies the space for researchers from different fields and backgrounds to translate and refine terminology and theory, all while working toward solving practical research problems. Consider, for example, research on curiosity. Studies with human children and adults often adopt an *information gap* approach (Kidd & Hayden, 2015), positing that people seek information to close a recognized

gap in their knowledge. By contrast, research with infants and nonhuman species (e.g., see Iwasaki & Kishimoto, 2021, for a study with primates) usually defines curiosity as a preference for novelty. In such scenarios, BTS can facilitate a comprehensive exploration of the target topic and contribute to developing broader, cross-taxa theoretical frameworks.

Improved Welfare Standards and Conservation Initiatives

BTS can assist research in adhering to ethical animal use principles (see Russell & Burch, 1959). First, BTS can help rationalize the number of animals whose testing is required to address a research question by providing a mechanism for sufficiently powered and better-coordinated studies. Second, large-scale collaborations can help maintain and improve housing and research standards by sharing best practices to encourage compliance with and consistency in ethical guidelines beyond institutional and geographical boundaries. This is especially beneficial for research conducted by nonmajor research or nonacademic institutions with limited resources to ensure the proper oversight of animal care and use. Finally, BTS can facilitate access to underrepresented or nonrepresented species. For instance, collaboration with zoos, aquariums, and sanctuaries can unlock access to threatened species that would otherwise remain unavailable for BTS groups, all while providing research opportunities for these institutions. For threatened species in particular, these data might inform important conservation actions (e.g., prerelease training; Greggor et al., 2014).

Conclusion

Comparative studies aim to achieve a comprehensive understanding of animal behavior and cognition, yet the narrow focus on a single taxon or species remains a heavy constraint. Here—as a group of researchers currently involved in the *ManyManys* collaboration—we make a call for BTS and argue that this approach can bring about noteworthy real-world advantages for comparative cognition research. These advantages include obtaining larger and more diverse samples, fostering best research practices, striking a balance between cross-species standardization and species-fair design, furthering theoretical advances, and improving welfare standards and conservation initiatives. By helping to overcome limitations in how research is conducted, BTS is uniquely poised to shape the future of comparative cognition for the better.

References

Bitterman, M. E. (1975). The comparative analysis of learning. *Science, 188*(4189), 699–709. https://doi.org/10.1126/science.188.4189.699

Castorina, J. A. (2021). The importance of worldviews for developmental psychology. *Human Arenas, 4*(2), 153–171. https://doi.org/10.1007/s42087-020-00115-9

Farrar, B. G., Boeckle, M., & Clayton, N. S. (2020). Replications in comparative cognition: What should we expect and how can we improve? *Animal Behavior and Cognition, 7*(1), 1–22. https://doi.org/10.26451/abc.07.01.02.2020

Farrar, B. G., Voudouris, K., & Clayton, N. S. (2021). Replications, comparisons, sampling and the problem of representativeness in animal cognition research. *Animal Behavior and Cognition, 8*(2), 273–295. https://doi.org/10.26451/abc.08.02.14.2021

Firestone, C. (2020). Performance vs. competence in human-machine comparisons. *Proceedings of the National Academy of Sciences, 117*(43), 26562–26571. https://doi.org/10.1073/pnas.1905334117

Forscher, P. S., Wagenmakers, E.-J., Coles, N. A., Silan, M. A., Dutra, N. B., Basnight-Brown, D., & IJzerman, H. (2023). The benefits, barriers, and risks of big-team science. *Perspectives on Psychological Science, 18*(3), 607–623. https://doi.org/10.1177/17456916221082970

Greggor, A. L., Clayton, N. S., Phalan, B., & Thornton, A. (2014). Comparative cognition for conservationists. *Trends in Ecology & Evolution*, *29*(9), 489–495. https://doi.org/10.1016/j.tree.2014.06.004

Iwasaki, S., & Kishimoto, R. (2021). Studies of prospective information-seeking in capuchin monkeys, pigeons, and human children. In J. R. Anderson & H. Kuroshima (Eds.), *Comparative cognition* (pp. 255–267). Springer. https://doi.org/10.1007/978-981-16-2028-7_15

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron, 88*(3), 449–460. https://doi.org/10.1016/j.neuron.2015.09.010

Lambert, M., Reber, S., Garcia-Pelegrin, E., Farrar, B., & Miller, R. (2022). ManyBirds: A multi-site collaborative approach to avian cognition and behaviour research. *Animal Behavior & Cognition, 9*(1), 133–152. https://doi.org/10.26451/abc.09.01.11.2022

Leavens, D. A., Bard, K. A., & Hopkins, W. D. (2010). BIZARRE chimpanzees do not represent "the chimpanzee." *Behavioral and Brain Sciences*, *33*(2–3), 100–101. https://doi.org/10.1017/S0140525X10000166

MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., Aureli, F., Baker, J. M., Bania, A. E., Barnard, A. M., Boogert, N. J., Brannon, E. M., Bray, E. E., Bray, J., Brent, L. J. N., Burkart, J. M., Call, J., Cantlon, J. F., Cheke, L. G., . . . Zhao, Y. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences*, *111*(20), E2140–E2148. https://doi.org/10.1073/pnas.1323533111

ManyBabies Consortium. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science, 3*(1), 24–52. https://doi.org/10.1177/2515245919900809

ManyDogs Project; Alberghina, D., Bray, E., Buchsbaum, D., Byosiere, S., Espinosa, J., Gnanadesikan, G. E., Alexandrina Guran, C.-N., Hare, E., Horschler, D. J., Huber, L., Kuhlmeier, V. A., MacLean, E. L., Pelgrim, M. H., Perez, B., Ravid-Schurr, D., Rothkoff, L., Sexton, C. L., Silver, Z. A., & Stevens, J. R. (2023). ManyDogs project: A Big Team Science approach to investigating canine behavior and cognition. *Comparative Cognition & Behavior Reviews, 18*, 59–77. https://doi.org/10.3819/CCBR.2023.180004

ManyPrimates; Altschul, D., Bohn, M., Canteloup, C., Ebel, S. J., Hanus, D., Hernandez-Aguilar, R. A., Joly, M., Keupp, S., Llorente, M., O'Madagain, C., Petkov, C. I., Proctor, D., Motes-Rodrigo, A., Sutherland, K., Szabelska, A., Taylor, D., Völter, C. J., & Wiggenhauser, N. G. (2022). Collaboration and open science initiatives in primate research. In B. L. Schwartz & M. J. Beran (Eds.), *Primate cognitive studies* (pp. 584–608). Cambridge University Press. https://doi.org/10.1017/9781108955836.023

Russell, W. M. S., & Burch, R. L. (1959). *The principles of humane experimental technique*. Methuen & Co.

Solvi, C., Baciadonna, L., & Chittka, L. (2016). Unexpected rewards induce dopamine-dependent positive emotion–like state changes in bumblebees. *Science, 353*(6307), 1529–1531. https://doi.org/10.1126/science.aaf4454

Tecwyn, E. C. (2021). Doing reliable research in comparative psychology: Challenges and proposals for improvement. *Journal of Comparative Psychology, 135*(3), 291–301. https://doi.org/10.1037/com0000291

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, *20*(4), 410–433. https://doi.org/10.1111/j.1439-0310.1963.tb01161.x

Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science, 13*(4), 411–417. https://doi.org/10.1177/1745691617751884

Voelkl, B., Vogt, L., Sena, E. S., & Würbel, H. (2018). Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLOS Biology*, *16*, Article e2003693. https://doi.org/10.1371/journal.pbio.2003693

Webster, M. M., & Rutz, C. (2020). How STRANGE are your study animals? *Nature, 582*(7812), 337–340. https://doi.org/10.1038/d41586-020-01751-5