Challenges and Promises of Big Team Comparative Cognition

Nicolás Alessandroni[1, *], Drew Altschul[2,3], Heidi A. Baumgartner[4], Marina Bazhydai[5], Sarah F. Brosnan[6,7], Krista Byers-Heinlein[1], Josep Call[8], Lars Chittka[9], Mahmoud Elsherif[10,11], Julia Espinosa[12], Marianne S. Freeman[13], Biljana Gjoneska[14], Onur Güntürkün[15], Ludwig Huber[16], Anastasia Krasheninnikova[17], Valeria Mazza[18,19], Rachael Miller[20,21], David Moreau[22], Christian Nawroth[23], Ekaterina Pronizius[24,25], Susana Ruiz-Fernández[26], Raoul Schwing[16], Vedrana Šlipogor[27,28,29], Ingmar Visser[30], Jennifer Vonk[31], Justin Yeager[32], Martin Zettersten[33,34], & Laurent Prétôt[35]

[1]Concordia University, Department of Psychology, Montréal, H4B 1R6, Canada

[2]The University of Edinburgh, School of Philosophy, Psychology and Language Sciences, Edinburgh, EH8 9JZ, UK

[3]Newcastle University, School of Psychology, Newcastle upon Tyne, NE2 4DR, UK

[4]Stanford University, Center for the Study of Language and Information, Stanford, CA 94305, USA

[5]Lancaster University, Psychology Department, Lancaster, LA1 4YF, UK

[6]Georgia State University, Departments of Psychology & Philosophy, Neuroscience Institute, Atlanta, GA 30303, USA

[7]National Center for Chimpanzee Care, MD Anderson Cancer Center, Bastrop, TX 78602, USA

[8]University of St Andrews, School of Psychology and Neuroscience, St Andrews, KY16 9JP, UK

[9]Queen Mary University of London, School of Biological & Behavioural Sciences, London, E1 4NS, UK

[10]University of Birmingham, School of Psychology, Birmingham, B15 2TT, UK

[11]University of Leicester, School of Psychology and Vision Sciences, Leicester, LE1 7RH, UK

[12]Harvard University, Department of Human Evolutionary Biology, Cambridge, MA 02138, USA

[13]University Centre Sparsholt, Animal Health and Welfare Research Centre, Winchester, SO21 2NF, UK

[14]Macedonian Academy of Sciences and Arts, Skopje, 1000, North Macedonia

[15]Ruhr University Bochum, Fakultät für Psychologie, Bochum, D-44780, Germany

[16]University of Veterinary Medicine Vienna, Messerli Research Institute, Vienna, 1210, Austria

[17]Max-Planck-Institute for Biological Intelligence, Department of Behavioural Neurobiology, Seewiesen, 82319, Germany

[18]University of Tuscia, Department for Ecological and Biological Sciences, Viterbo, 01100, Italy

[19]University of Potsdam, Animal Ecology Group, Postdam, 14469, Germany

[20]Anglia Ruskin University, School of Life Sciences, Cambridge, CB1 1PT, UK

[21]University of Cambridge, Department of Psychology, Cambridge, CB2 3EB, UK

[22]University of Auckland, School of Psychology and Centre for Brain Research, Auckland, 1010, New Zealand

[23]Research Institute for Farm Animal Biology, Working Group Animal Behaviour & Welfare, Dummerstorf, 18196, T

[24]University of Vienna, Department of Cognition, Emotion, and Methods in Psychology, Vienna, 1010, Austria

[25]UCLouvain, Psychological Sciences Research Institute, Louvain-la-Neuve, 1348, Belgium

[26]Brandenburg University of Technology Cottbus – Senftenberg, Department of Psychology, Cottbus, 03048, Germany

[27]University of Lausanne, Department of Ecology and Evolution, Lausanne, CH-1015, Switzerland

[28]The Sense Innovation and Research Center, Lausanne & Sion, Lausanne, CH-1007, Switzerland

[29]University of South Bohemia, Department of Zoology, České Budějovice, 370 05, Czech Republic

[30]University of Amsterdam, Department of Psychology, Amsterdam, 1001 NK, The Netherlands

[31]Oakland University, Department of Psychology, Rochester, MI 48309-4482, USA

[32]Universidad de las Américas, Grupo de Investigación en Biodiversidad, Medio Ambiente y Salud (BIOMAS), Facultad de Ingenierías y Ciencas Aplicadas, Quito, 170124, Ecuador

[33]University of California San Diego, Department of Cognitive Science, La Jolla, CA 92093, USA

[34]Princeton University, Department of Psychology, Princeton, NJ 08540, USA

[35]Pittsburg State University, Department of Psychology and Counseling, Pittsburg, KS 66762, USA

*Corresponding author: Nicolás Alessandroni (nicolas.alessandroni@concordia.ca).

Standfirst

Big Team Science (BTS) has the potential to reshape comparative cognition research, but its implementation—especially making comparisons species-fair, handling multi-site variation, and reaching researcher consensus—poses daunting challenges. Here, we propose solutions and discuss how BTS can transform the field.

Big Team Science (BTS) is a powerful approach to scientific discovery wherein teams of researchers pool resources and efforts to answer fundamental questions in their fields. BTS offers clear advantages, most notably the ability to achieve greater sample size and diversity than any individual or team, which further improves the reliability and generalizability of inferences. While BTS has long been the norm in the natural sciences such as astronomy, physics and genetics, successful BTS collaborations have emerged only recently in the behavioral, cognitive, and social sciences, spanning both single species (e.g., ManyBabies, ManyDogs) and multiple species (e.g., ManyPrimates, ManyBirds). BTS across *multiple* and *diverse* taxa is an even more recent development (e.g., ManyManys). We recently argued that BTS offers unique advantages to the study of comparative cognition, including balancing cross-species standardization and species-fair designs, advancing theories, and improving welfare standards and conservation initiatives[1]. Yet, implementing BTS poses challenges, some of which are unique to cross-species comparisons. Using our experience as members of the *ManyManys* collaboration (https://manymanys.github.io), we highlight three of those challenges, together with potential solutions.

Implementing Species-Fair Comparisons

Comparative cognition aims to elucidate ontogenetic and phylogenetic differences across species. This ambitious endeavor improves our understanding of individual animal models while shedding light on human behavior and development[2]. Unlike disciplines that focus on a single species, comparative cognition must grapple with organisms with distinct sensory modalities, morphological and physiological adaptations, and behavioral repertoires, which calls for complex methodological solutions. One promising solution to address this challenge lies in the so-called *species-fair approach,* which argues that successful cross-species comparisons hinge on a delicate balance between standardizing core task components and permitting procedural modifications aligned with species-specific needs and preferences[3].

Admittedly, in the context of BTS, implementing a species-fair approach can be challenging when there is a large number and diversity of taxa involved. In this case, task design, experimental features (e.g., task parameters), and species-specific characteristics (e.g., sensory modalities) require even more careful consideration than in single-species/single-lab studies to ensure construct validity. For example, a task comparing memory abilities across different species might need to include different types of cues (e.g., visual, auditory, olfactory) to accommodate species' differences in sensory abilities. Put differently, diverse skills or traits necessitate tailoring tasks to ensure that findings accurately reflect the underlying processes of the species under study. Similarly, species-specific properties might lead to different optimal values for task parameters. For instance, a 15-minute intertrial interval might be reasonable when testing fish in an aquarium, but unsuitable for assessing monkeys in a testing box[4]. When these differences are not considered carefully, failures to successfully complete a task might be due to inappropriate methodological design rather than true cognitive differences.

Not only must standardized tasks cater to the specific needs of diverse species, but they must be ecologically relevant and hold motivational value for the species tested. For example, infants can be rewarded with visually attractive pictures and/or objects, while other species may respond to food or access to conspecifics as effective reinforcers. Importantly, even if species share a particular ability, they might display it in distinct contexts or it may be underpinned by different mechanisms, which calls for discretion before drawing conclusions about the observed behaviors. For instance, different species (or even individuals of the same species) might pass the same test using entirely different behavioral strategies (e.g. prosocial vs. spiteful)[5]. Therefore, it is vital to establish whether a species can solve a task, quantify its performance, and recognize the behavioral signatures of the animal (i.e., strategies, types of errors, or biases).

Fortunately, standardized methods for implementing species-fair designs are being developed. For example, some scholars advocate for a two-pronged approach where subjects should first be compared using tasks in which parameters are as identical as possible to establish a performance baseline across species[4], after which they should be compared on tasks adapted to their particular characteristics, needs, and preferences[6]. This approach allows researchers to both compare species and explore the impact of species-specific factors using tasks that are optimized for each species. Due to their increased sample size and diversity, BTS projects provide an excellent platform for empirically testing how different species-fair approaches actually perform.

Moreover, in a BTS context, researchers benefit from a distributed knowledge system that makes planning such testing more feasible. Unlike more traditional research approaches for which one or a few researcher(s) is/are expected to possess extensive knowledge across multiple species, BTS capitalizes on the collective expertise of a diverse team, as well as access to more diverse species and populations. Each member contributes with a specialized understanding of different species, and this wider array of perspectives and insights enriches the research process and aids in the interpretation of results. Such a collaborative structure facilitates more nuanced and species-appropriate experimental designs.

Measuring and Controlling for Multi-Site Differences

In comparative cognition, BTS collaborations typically collect data from several testing sites. This broadens sample sizes, but can also introduce substantial cross-site variability that obscures true differences between groups or species. For example, intraspecific cognitive variation can arise due to differences in rearing histories between (and potentially within) sites[7]. Additionally, factors such as subjects' testing experience and experimenter bias can facilitate or hinder performance on a task[8,9]. Moreover, differences across sites can also result from limitations in controlling or manipulating key variables, including those associated with food and the separation of subjects[10]. These limitations are especially prevalent in zoos and other non-laboratory settings, which are common research environments within comparative cognition research. Additional layers of complexity are added when considering current efforts towards studying cognition directly in the wild and/or under ecologically relevant conditions[11], or when testing human children in non-experimental environments beyond the experimenter's control[12].

For example, when two or more sites fail to replicate findings, it is often assumed that the differences are due to experimental errors, or that results are contradictory or even invalid. However, it can also be the case that the observed differences reveal the true phenotypic variability expressed by different populations of the same species under different conditions. In such a scenario, BTS emerges as a unique opportunity to disentangle local patterns from generalizable processes and to provide a better estimate of a species' phenotypic variability. By incorporating multi-site data, BTS enables the mapping of variability that supports a better understanding of the full scope of species' cognitive repertoires.

Concrete strategies can help BTS researchers understand site- and experimenter-specific factors. For example, researchers can document details on design implementation and share precise and detailed information on possible inconsistencies. In ManyBabies 1—a large-scale collaboration that replicated the finding that infants prefer infant-directed speech over adult-directed speech—each participating lab documented its setup and experimental procedure by submitting walkthrough videos, thus providing rich information about each step of the data collection procedure and how it might have varied across labs[13]. In ManyPrimates 1, the lead team collected information on species and site-relevant factors—such as the size of the experimental apparatus and social group size—as well as individual factors—including prior task experience—and integrated them into phylogenetic analyses[14]. Additionally, automated setups, data acquisition protocols, and closed-loop settings in which animals interact with computer-controlled setups help insulate data acquisition against observer bias as well as unintended cues from human experimenters[15]. Future directions for behavioral research might include exploring machine learning and AI-powered tools to facilitate objective data acquisition and improve the precision and comparability of behavioral data across study sites.

Navigating Decision-Making and Theoretical Divides

All BTS networks must agree on crucial aspects of project implementation, including topic selection, conceptual definitions, and study design. However, group decision-making in comparative cognition research presents particular challenges due to the diverse theoretical and disciplinary backgrounds of the participants. The core risk lies in the potential for a narrow range of perspectives—often from already overrepresented groups—to dictate research questions, theories, and methodologies. Consider a project exploring learning skills across species. In such a project, biologists and psychologists confront the challenge of reconciling their diverse, discipline-specific definitions. When deciding on the research methodology, there could be disagreements on whether to prioritize observations of natural behaviors in the wild or controlled experiments that manipulate environmental variables. This divergence across fields can lead to difficulties in agreeing on what constitutes evidence for learning, how to measure it, and even which species to study.

To facilitate meaningful comparisons across taxa, researchers need to use common terminology to refer to cognitive capacities across species and clearly define the criteria for interpreting behaviors. For instance, in the competition for high-profile publications about "clever" animals, definitions are sometimes adjusted so that the favored species "pass" the test. However, the more permissive the definition, the more species will qualify, and the more examples of functional convergence (or indeed homology) one might (mis-)identify. In other words, there is a risk that identifying evolutionary patterns in cognition depends somewhat on semantics rather than actual biological traits[14]. The challenge in BTS is twofold: aligning terminologies and reconciling different theoretical accounts and methodologies. The successful integration of diverse perspectives requires an in-depth understanding and appreciation of each discipline's theoretical underpinnings and a commitment to developing hybrid frameworks to accommodate discrepancies.

To help solve these challenges, BTS networks need to implement strategies for interdisciplinary integration. For example, BTS projects can take a *consensus-building approach*, aiming to make decisions that are generally agreed upon by as many team members as possible. Consensus-building activities can include (a) soliciting and facilitating open discussions in targeted workshops or meetings, (b) providing multiple forums and mechanisms for collecting feedback on team decisions, and (c) polling opinions from the entire research team on key decision points in the project. In comparative cognition, it is particularly important that consensus is developed in a democratic manner with input from researchers representing a wide range of perspectives and taxa.

Conclusion

Implementing a BTS approach presents hurdles for research projects in all fields, with comparative cognition facing particularly substantial challenges. We have argued that there are promising opportunities to refine research methodologies and workflows to overcome these challenges. BTS holds immense potential to foster interdisciplinary integration, facilitate group decision making, and advance open science through global partnerships. By leveraging the diverse expertise of scientists in task design, data interpretation, and the comprehensive mapping of species' phenotypic variability using multi-site data, BTS can reshape research practices in comparative cognition and accelerate significant advancements in the field.

References

1. Alessandroni, N., Altschul, D., Bazhydai, M., Byers-Heinlein, K., Elsherif, M. M., Gjoneska, B., Huber, L., Mazza, V., Miller, R., Nawroth, C., Pronizius, E., Qadri, M. A. J., Šlipogor, V., Soderstrom, M., Stevens, J. R., Visser, I., Williams, M., Zettersten, M., & Prétôt, L. (2024). Comparative cognition needs Big Team Science: How large-scale collaborations will unlock the future of the field. *Comparative Cognition & Behavior Reviews*, *19*, 67-72.
2. Maestripieri, D. (2005). On the importance of comparative research for the understanding of human behavior and development: A reply to Gottlieb & Lickliter (2004). *Social Development*, *14*(1), 181-186.
3. 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.
4. Salwiczek, L. H., Prétôt, L., Demarta, L., Proctor, D., Essler, J., Pinto, A. I., Wismer, S., Stoinski, T., Brosnan, S. F., & Bshary, R. (2012). Adult cleaner wrasse outperform capuchin monkeys, chimpanzees and orang-utans in a complex foraging task derived from cleaner–client reef fish cooperation. *PLoS One*, *7*(11), e49068.
5. Tennie, C., Jensen, K. & Call, J. (2016). The nature of prosociality in chimpanzees. *Nature Communications, 7*, 13915.
6. Prétôt, L., Bshary, R., & Brosnan, S. F. (2016). Factors influencing the different performance of fish and primates on a dichotomous choice task. *Animal Behaviour, 119,* 189-199.

7. Lyn, H., Russell, J. L., & Hopkins, W. D. (2010). The impact of environment on the comprehension of declarative communication in apes. *Psychological Science, 21*(3), 360-365.

8. Prétôt, L., Mickelberg, J., Carrigan, J., Stoinski, T., Bshary, R., & Brosnan, S. F. (2021). Comparative performance of orangutans (*Pongo spp.*), gorillas (*Gorilla gorilla gorilla*), and drills (*Mandrillus leucophaeus*), in an ephemeral foraging task. *American Journal of Primatology*, *83*(1), e23212.

9. Clark, H., Elsherif, M. M., & Leavens, D. A. (2019). Ontogeny vs. Phylogeny in primate/canid comparisons: A meta-analysis of the object choice task. *Neuroscience and Biobehavioral Reviews, 105*, 178–189.

10. Talbot, C. F., Parrish, A. E., Watzek, J., Essler, J. L., Leverett, K. L., Paukner, A., & Brosnan, S. F. (2018). The influence of reward quality and quantity and spatial proximity on the responses to inequity and contrast in capuchin monkeys (*Cebus [Sapajus] apella*). *Journal of Comparative Psychology, 132*(1), 75-87.

11. Morand-Ferron, J., Cole, E. F., & Quinn, J. L. (2016). Studying the evolutionary ecology of cognition in the wild: A review of practical and conceptual challenges. *Biological Reviews*, *91*(2), 367-389.

12. Tamis-LeMonda, C. S., & Masek, L. R. (2023). Embodied and embedded learning: Child, caregiver, and context. *Current Directions in Psychological Science, 32*(5), 369-378.

13. The 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.

14. ManyPrimates, Aguenounoun, G., Allritz, M., Altschul, D. M., Ballesta, S. Beaud, A., Bohn, M., Bornbusch, S. L., Brandão, A., Brooks, J., Bugnyar, T., Bustamante, L., Call, J., Canteloup, C., Caspar, K. R., de Sousa, A. A., DeTroy, S. E., Eppley, T. M., Fichtel, C., … & Zablocki-Thomas, P. (2022). The evolution of primate short-term memory. *Animal Behavior and Cognition, 9*(4), 428-516.

15. Chittka, L., Rossiter, S. J., Skorupski, P., & Fernando, C. (2012). What is comparable in comparative cognition? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2677-2685.

Competing interests

The authors declare no conflicts of interest.

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 BTS collaboration. The authors are deeply grateful to Hilary Killam for creating the image featured in the manuscript. This work was supported by the following fellowships and grants: FRQSC Postdoctoral Fellowship (B3Z, #333109) [NA], British Academy PF20\100086 [DA], NSF BCS 2127375 [SFB], Open Philanthropy [LC], Leverhulme Early-Career Research Fellowship [ME], National Science Foundation Award #2209046 [JE], ERC-2020-ADG, AVIAN MIND, LS5, GA No. 101021354 [OG], The Sense - Innovation and Research Center Lausanne & Sion (#CFP2023_WLDC), USB Postdoctoral Fellowship [VŠ], UDLA Grant 483.A.XIV.24 (JY), NIH NICHD F32HD110174 [MZ], K-INBRE P20 GM103418 [LP].

Figure captions

Accompanying Figure: Illustration by Hilary Killam. Published under a CC-BY-NC 4.0 license. Inspired by the image "Community exchange," created by Scriberia with The Turing Way community [CC-BY 4.0 license, available at https://doi.org/10.5281/zenodo.13882307].