https://doi.org/10.1038/s41562-024-02081-6

Challenges and promises of big team comparative cognition

Nicolás Alessandroni, Drew Altschul, Heidi A. Baumgartner, Marina Bazhydai, Sarah F. Brosnan, Krista Byers-Heinlein, Josep Call, Lars Chittka, Mahmoud Elsherif, Julia Espinosa, Marianne S. Freeman, Biljana Gjoneska, Onur Güntürkün, Ludwig Huber, Anastasia Krasheninnikova, Valeria Mazza, Rachael Miller, David Moreau, Christian Nawroth, Ekaterina Pronizius, Susana Ruiz-Fernández, Raoul Schwing, Vedrana Šlipogor, Ingmar Visser, Jennifer Vonk, Justin Yeager, Martin Zettersten & Laurent Prétôt

Big team science has the potential to reshape comparative cognition research, but its implementation – especially in making fair comparisons between species, handling multisite variation and reaching researcher consensus – poses daunting challenges. Here, we propose solutions and discuss how big team science can transform the field.

Big team comparative cognition; adapted with permission from an illustration by Hilary Killam (inspired by the image 'Community exchange', created by Scriberia with The Turing Way community; available at https://doi.org/10.5281/zenodo.6821117).

Big team science (BTS) is a powerful approach to scientific discovery, in which 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.

Although 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 behavioural, cognitive and social sciences, and span both single (for example, ManyBabies or ManyDogs) and multiple (for example, ManyPrimates or ManyBirds) species. BTS across multiple and diverse taxa is an even more recent development (for example, ManyManys). Elsewhere, we 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, we highlight three of these challenges, together with potential solutions.

Implementing 'species-fair' comparisons

Comparative cognition aims to elucidate ontogenetic and phylogenetic differences across species. This ambitious endeavour improves our understanding of individual animal models, while shedding light on human behaviour 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 behavioural repertoires, which calls for complex methodological solutions. One promising solution to addressing 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 that are 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 (for example, task parameters) and species-specific characteristics (for example, sensory modalities) require even more careful consideration than in single-species or single-laboratory studies to ensure construct validity. For example, a task that compares memory abilities across different species might need to include different types of cues (for example, visual, auditory or 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-min 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 also be ecologically relevant and hold motivational value for the species tested. For example, infants can be rewarded with visually attractive pictures and/or objects, whereas 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 behaviours. For instance, different species (or even individuals of the same species) might pass the same test using entirely different behavioural strategies (for example, prosocial versus spiteful)[5]. Therefore, it is vital to establish whether a species can solve a task, quantify its performance and recognize the behavioural signatures of the animal (that is, 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 in which participants or 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 enables researchers to both compare species and explore the effect of species-specific factors using tasks that are optimized for each species. Owing 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 researcher is (or a few researchers 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 owing to differences in rearing histories between (and potentially within) sites[7]. Additionally, factors such as participants' or 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 participants or 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 nonexperimental 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 multisite 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 to understand site-specific 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 laboratory documented its setup and experimental procedure by submitting walkthrough videos, and thus provided rich information about each step of the data collection procedure and how it might have varied across laboratories[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 to insulate data acquisition against observer bias as well as unintended cues from human experimenters[15]. Future directions for behavioural research might include exploring machine learning and artificial intelligence-powered tools to facilitate objective data acquisition and improve the precision and comparability of behavioural 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 owing 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 that explores 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 behaviours 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 behaviours. For instance, in the competition for high-profile publications about 'clever' animals, definitions are sometimes adjusted so that the favoured 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 identify or misidentify. 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 to solve these challenges, BTS networks need to implement strategies for interdisciplinary integration. For example, BTS projects can take a consensus-building approach and aim 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, and comparative cognition faces 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 multisite data, BTS can reshape research practices in comparative cognition and accelerate important advancements in the field.

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]Department of Psychology, Concordia University, Montréal, Québec, Canada. [2]School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, Edinburgh, UK. [3]School of Psychology, Newcastle University, Newcastle upon Tyne, UK. [4]Center for the Study of Language and Information, Stanford University, Stanford, CA, USA. [5]Psychology Department, Lancaster University, Lancaster, UK. [6]Neuroscience Institute, Departments of Psychology & Philosophy, Georgia State University, Atlanta, GA, USA. [7]National Center for Chimpanzee Care, MD Anderson Cancer Center, Bastrop, TX, USA. [8]School of Psychology and Neuroscience, University of St Andrews, St Andrews, UK. [9]School of Biological & Behavioural Sciences, Queen Mary University of London, London, UK. [10]School of Psychology, University of Birmingham, Birmingham, UK. [11]School of Psychology and Vision Sciences, University of Leicester, Leicester, UK. [12]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA. [13]Animal Health and Welfare Research Centre, University Centre Sparsholt, Winchester, UK. [14]Macedonian Academy of Sciences and Arts, Skopje, North Macedonia. [15]Fakultät für Psychologie, Ruhr University Bochum, Bochum, Germany. [16]Messerli Research Institute, University of Veterinary Medicine Vienna, Vienna, Austria. [17]Department of Behavioural Neurobiology, Max Planck Institute for Biological Intelligence, Seewiesen, Germany. [18]Department for Ecological and Biological Sciences, University of Tuscia, Viterbo, Italy. [19]Animal Ecology Group, University of Potsdam, Postdam, Germany. [20]School of Life Sciences, Anglia Ruskin University, Cambridge, UK. [21]Department of Psychology, University of Cambridge, Cambridge, UK. [22]Centre for Brain Research, School of Psychology, University of Auckland, Auckland, New Zealand. [23]Working Group Animal Behaviour & Welfare, Research Institute for Farm Animal Biology, Dummerstorf, Germany. [24]Department of Cognition, Emotion, and Methods in Psychology, University of Vienna, Vienna, Austria. [25]Psychological Sciences Research Institute, UCLouvain, Louvain-la-Neuve, Belgium. [26]Department of Psychology, Brandenburg University of Technology Cottbus – Senftenberg, Cottbus, Germany. [27]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland. [28]The Sense Innovation and Research Center, Lausanne and Sion, Lausanne, Switzerland. [29]Department of Zoology, University of South Bohemia, České Budějovice, Czech Republic. [30]Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands. [31]Department of Psychology, Oakland University, Rochester, MI, USA. [32]Grupo de Investigación en Biodiversidad, Medio Ambiente y Salud (BIOMAS), Facultad de Ingenierías y Ciencas Aplicadas, Universidad de las Américas, Quito, Ecuador. [33]Department of Cognitive Science, University of California San Diego, La Jolla, CA, USA. [34]Department of Psychology, Princeton University, Princeton, NJ, USA. [35]Department of Psychology and Counseling, Pittsburg State University, Pittsburg, KS, USA.

✉e-mail: nicolas.alessandroni@concordia.ca

Published online: 18 December 2024

References

1. Alessandroni, N. et al. *Comp. Cogn. Behav. Rev.* **19**, 67–72 (2024).
2. Maestripieri, D. *Soc. Dev.* **14**, 181–186 (2005).
3. Farrar, B. G., Voudouris, K. & Clayton, N. S. *Anim. Behav. Cogn.* **8**, 273–295 (2021).
4. Salwiczek, L. H. et al. *PLoS One* **7**, e49068 (2012).
5. Tennie, C., Jensen, K. & Call, J. *Nat. Commun.* **7**, 13915 (2016).
6. Prétôt, L., Bshary, R. & Brosnan, S. F. *Anim. Behav.* **119**, 189–199 (2016).
7. Lyn, H., Russell, J. L. & Hopkins, W. D. *Psychol. Sci.* **21**, 360–365 (2010).
8. Prétôt, L. et al. *Am. J. Primatol.* **83**, e23212 (2021).
9. Clark, H., Elsherif, M. M. & Leavens, D. A. *Neurosci. Biobehav. Rev.* **105**, 178–189 (2019).
10. Talbot, C. F. et al. *J. Comp. Psychol.* **132**, 75–87 (2018).
11. Morand-Ferron, J., Cole, E. F. & Quinn, J. L. *Biol. Rev. Camb. Philos. Soc.* **91**, 367–389 (2016).
12. Tamis-LeMonda, C. S. & Masek, L. R. *Curr. Dir. Psychol. Sci.* **32**, 369–378 (2023).
13. The ManyBabies Consortium. *Adv. Methods Pract. Psychol. Sci.* **3**, 24–52 (2020).
14. ManyPrimates et al. *Anim. Behav. Cogn.* **9**, 428–516 (2022).
15. Chittka, L., Rossiter, S. J., Skorupski, P. & Fernando, C. *Phil. Trans. R. Soc. Lond. B* **367**, 2677–2685 (2012).

Acknowledgements

The authors of this work are listed alphabetically with the exception of N.A. (first and corresponding author) and L.P. (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 H. Killam for creating the original version of the image featured in the manuscript. The authors were supported by the following fellowships and grants: FRQSC Postdoctoral Fellowship (B3Z, no. 333109) (N.A.), British Academy PF20\100086 (D.A.), NSF BCS 2127375 (S.F.B.), Open Philanthropy (L.C.), Leverhulme Early-Career Research Fellowship (M.E.), National Science Foundation Award no. 2209046 (J.E.), ERC-2020-ADG, AVIAN MIND, LS5, GA no. 101021354 (O.G.), National Biodiversity Future Center—NBFC, "NextGenerationEU" (Mission 4 Component 2 Investment 1.4—Project code CN_00000033, CUP J83C22000860007) (V.M.), The Sense Innovation and Research Center, Lausanne and Sion (no. CFP2023_WLDC), USB Postdoctoral Fellowship (V.Š.), UDLA Grant 483.A.XIV.24 (J.Y.), NIH NICHD F32HD110174 (M.Z.) and K-INBRE P20 GM103418 (L.P.).

Competing interests

The authors declare no competing interests.