Trends in Cognitive Sciences

CellPress

Scientific Life

Generative adversarial collaborations: a new model of scientific discourse

Benjamin Peters[1,10], Gunnar Blohm[2,10], Ralf Haefner[3], Leyla Isik[4], Nikolaus Kriegeskorte[5], Jennifer S. Lieberman[5], Carlos R. Ponce[6], Gemma Roig[7], and Megan A.K. Peters[8,9,*]

Science progresses when ideas clash, leaving the most successful to survive and move us closer to the truth. In this ideal hypothetico-deductive approach [1], science is dynamic and fluid, with theories constantly tested and replaced. In reality, however, many opposing theories rarely meet. Scientists instead often work in entrenched paradigms or research programs – focused on their own frameworks, language, and methods – which resist direct comparison and evolve incrementally at a generational timescale rather than through confrontations [2,3]. Adversarial collaborations offer a promising alternative to accelerate scientific progress: a way to bring together researchers from different camps to rigorously compare and test their competing views [4,5].

Cognitive computational neuroscience is an emerging interdisciplinary field with a dedicated conference [6] and community-driven events. In 2020, the Cognitive Computational Neuroscience (CCN) conference (https://www.ccneuro.org) launched the Generative Adversarial Collaboration (GAC) project (https://gac.ccneuro.org/), aiming to reveal, discuss, and advance some of the most tantalizing controversies in the field. Over the past 5 years this program has organized 17 GACs with over 100 different scientists and recorded kickoff events at the CCN conference, leading to preprints and peer-reviewed publications of key conceptual pieces covering fundamental questions of the human brain and mind (Box 1). We believe that the GAC model may be successful in other scientific disciplines. In this perspective we, therefore, describe our motivations for and experiences of organizing the GAC project in hopes that our experiences will be valuable to those who wish to implement a similar program.

Benefits and challenges of adversarial collaborations

Across various areas of research, adversarial collaborations [7,8] promise conceptual and practical benefits for researchers, the scientific community, and society by accelerating scientific progress. Here we define an 'adversarial collaboration' as one in which researchers who espouse different positions on a particular issue collaborate with the goal of making progress towards a resolution [9].

Recent instances of adversarial collaborations [8,10,11] have focused on empirical theory arbitration, that is, defining and running a set of 'crucial experiments' [12–14] to adjudicate among theories, such that one theory concedes defeat while another survives. In practice, however, entire theories are rarely disproven or 'falsified' by a few experiments. Instead, when encountering contradicting evidence, scientific theories adapt to account for new empirical data rather than being dismissed by their proponents [2,3]. Hence, adversarial collaborations often conclude with all involved theories still supported by their proponents, though some may be forced to refine their positions based on crucial experiments' results [8].

But even if adversarial collaborations conclude without one side conceding defeat, they are still highly valuable. First, a published report on the adversarial collaboration serves as a signpost, allowing the scientific community to make their own judgments about each theory's merits. Second, even without conducting crucial experiments, adversarial collaborations can result in better scientific communication. For example, a frequent cause of apparent theoretical disagreement can result from simple misunderstandings or conceptual or terminological differences that reduce to semantics; adversarial collaborations can eliminate misunderstandings and find agreement simply through the process of discussion in good faith, often revealing more commonalities between theories than expected. Adversarial collaborations therefore first need to develop a common language and methodological agreement as a basis for designing crucial experiments and communicating their outcomes. We believe that a common language, agreed-upon methods, and mutual conceptual understanding are the outcomes of adversarial collaborations that have the greatest positive impact. The CCN GAC project therefore aims to bring these benefits to the forefront, focusing on transparency at all stages of the adversarial collaboration process rather than just the later stages of crucial experiments and empirical results.

The process of GACs at CCN

GACs aim to break intellectual silos of opposing theoretical camps, leading to a common language, agreed-upon methods, and a mutual understanding of opposing views. GACs thus aim to be generative, inspiring novel research programs by strongly emphasizing transparency and community engagement, most often culminating in the output of a position piece rather than requiring 'theory-killing' experimental outcomes or acquisition of grant funding. This goal is dictated also by the needs of cognitive computational neuroscience – a relatively young and highly interdisciplinary field, which emerged with its own conference only a

Box 1. CCN GAC quick facts

- Year of launch: 2020
- GACs to date: 17
- Unique scientists involved: 118+
- Position papers: 8
- Proposals, workshop recordings, and paper links: https://gac.ccneuro.org/gacs-by-year

few years ago [6]. The field is changing quickly, continually incorporating new advances from artificial intelligence, neuroengineering, computational neuroscience, and cognitive science. Productive disagreements and debates can emerge in this context when highly different conceptual approaches encounter each other. Thus, theory falsification, championed as a strength of adversarial collaborations, is not the only kind of desirable outcome. Before we aim to empirically arbitrate and 'kill' theories in our young field, we want them to mature through discussion, discovery, and community education.

These needs drive the 'generative' component of the GACs, with emphasis on transparency at every stage (Figure 1, top). From the moment of team formation, teams are encouraged to involve researchers from multiple backgrounds and career stages. Teams then develop and submit a GAC proposal laying out the controversy, its importanceto the CCN community, and the team itself. These proposals are reviewed by the CCN GAC organizer subcommittee for clarity and relevance, then posted publicly for open review and to create the opportunity for teams to welcome new members. An interactive kickoff workshop is then hosted as a center-stage event at the CCN conference: competing theories are presented and debated (with emphasis on viewpoint integration), audience discussion is encouraged, and teams may further expand. Post-workshop, GAC teams continue discussions, culminating in a collaborative position paper published in an ongoing special issue in the *Neurons, Behavior, Data, and Theory* journal. The position papers aim to coherently describe the controversy's evolution throughout the GAC process – including the common language and shared assumptions and interpretations that have been established, resolutions or alignment, and remaining controversies – and invite community input. These papers can then

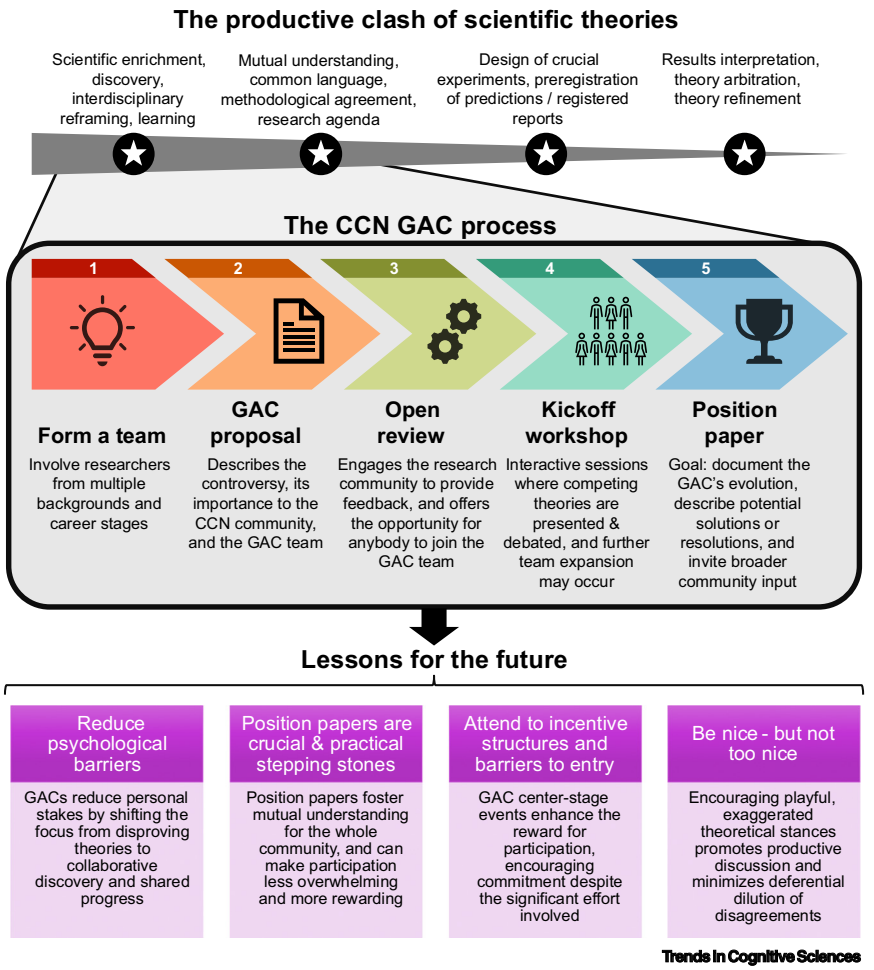

Figure 1. Overview of the Cognitive Computational Neuroscience (CCN) Generative Adversarial Collaboration (GAC) process. Focused on the earlier stages of an adversarial collaboration – before the design, execution, and interpretation of critical experiments – every stage of the GAC process is designed to maximize transparency, engagement, and community benefit. The end result is not necessarily a set of crucial experiments; instead, GACs typically 'end' with a position paper published in the *Neurons, Behavior, Data Analysis and Theory* journal laying out the controversy, its evolution throughout the GAC process, and any resolutions or conceptual alignment that may have taken place. Through organizing the CCN GACs, we have learned the value of reducing psychological barriers to engagement, providing support to develop and share position papers, attending to potential barriers to participation and success, and encouraging active debate.

inspire others even outside the GAC team to engage in novel research, serving to outline a roadmap for empirical work that goes beyond the members of a particular GAC.

Lessons learned

We have learned much through organizing the CCN GAC project [15]. We summarize these lessons' themes here (Figure 1, bottom), in hopes they will be useful to organizers of other adversarial collaborations.

Reduce psychological barriers

Researchers may be reluctant to participate in adversarial collaborations because they are personally invested in their work and fear disproving their own theories. The CCN GAC program addresses this by reframing the goal away from only (adversarially) challenging theories towards focusing on (generative) interdisciplinary exchange, discovery, and theory evolution. By emphasizing the shared ownership of the process, involving the community, and championing the intrinsic joy of jointly discovering uncharted territory between opposing positions, the process becomes less threatening, allowing researchers to gain a stake in the collective outcome without feeling their theories are at risk.

Position papers are crucial and practical stepping stones

Adversarial collaborations require significant time investment, and the scope of devising and conducting crucial experiments can overwhelm and deter participation. To counter this, the CCN GAC program aims to yield position papers as deliverables. We believe that position papers using common language and emphasizing mutual understanding of opposing positions, are the most valuable outcome of an adversarial collaboration, as they inspire novel research and lead to theory enrichment and refinement. Practically, position papers as target outcomes also relieve some burden on participants and incentivize commitment by offering opportunities for citation.

Attend to incentive structures and barriers to entry

An adversarial collaboration may have too-small incentives and an imbalance between effort and reward, deterring participation. We have found that the elevated profile as center-stage open community events at CCN can alleviate this challenge through offering immediate, highly visible recognition of GAC member contributions – an important consideration especially for early-career researchers who can often drive a GAC forward.

Be nice – but not too nice

Culture and social norms may lead scientists to avoid open conflict – even after committing to GAC participation. These norms favor language that underplays disagreements and hedges theoretical positions, which could lead a GAC position paper to simply offer an intricate tessellation of mostly-unchanged theoretical positions rather than reflecting the outcome of productive discussion or disagreement. We have found that encouraging GAC members to playfully take on more accentuated theoretical stances in service of debate, and including or dedicating contrarian roles to members, can contribute to the quality and impact of the GAC.

Closing thoughts

The CCN GAC initiative aims to maximize the positive impact of the adversarial collaboration model by dismantling intellectual isolation through transparent, community-engaged dialogue and emphasis on process in addition to final results. Due to the GAC project's impact (Box 1) in this young and interdisciplinary field – where perspectives, backgrounds, and approaches differ between contributing disciplines – we believe the CCN GAC model may accelerate and enhance scientific discourse also in other fields where discourse remains largely sequestered behind closed doors.

Acknowledgments

The authors thank the Cognitive Computational Neuroscience conference community for their support of and participation in the Generative Adversarial Collaboration project. The authors have no funding sources to declare.

[1]School of Informatics, University of Edinburgh, Edinburgh, UK
[2]Centre for Neuroscience Studies, Queen's University, Kingston, Ontario, Canada
[3]Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA
[4]Department of Cognitive Science, Johns Hopkins University, Baltimore, MD, USA
[5]Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA
[6]Department of Neurobiology, Harvard Medical School, Boston, MA, USA
[7]Department of Computer Science, Goethe Frankfurt University, Frankfurt a.M., Germany
[8]Department of Cognitive Sciences, University of California Irvine, Irvine, CA, USA
[9]Program in Brain, Mind, and Consciousness, Canadian Institute of Advanced Research, Toronto, Ontario, Canada
[10]These authors share first authorship.

*Correspondence:
megan.peters@uci.edu (M.A.K. Peters).

https://doi.org/10.1016/j.tics.2024.10.015

References

1. Popper, K.R. (1959) *The logic of scientific discovery*, Basic Books, Oxford, England
2. Kuhn, T.S. (1962) *The structure of scientific revolutions*, University of Chicago Press, Chicago
3. Lakatos, I. (1968) Criticism and the methodology of scientific research programmes. *Proc. Aristot. Soc.* 69, 149–186
4. Kahneman, D. (2003) Experiences of collaborative research. *Am. Psychol.* 58, 723–730
5. Latham, G.P. *et al.* (1988) Resolving scientific disputes by the joint design of crucial experiments by the antagonists: Application to the Erez–Latham dispute regarding participation in goal setting. *J. Appl. Psychol.* 73, 753–772
6. Naselaris, T. *et al.* (2018) Cognitive Computational Neuroscience: A New Conference for an Emerging Discipline. *Trends Cogn. Sci.* 22, 365–367
7. Ceci, S.J. *et al.* (2024) Adversarial collaboration: An undervalued approach in behavioral science. *Am. Psychol.*, Published online August 15, 2024. https://doi.org/10.1037/amp0001391
8. Cowan, N. *et al.* (2020) How do scientific views change? Notes from an extended adversarial collaboration. *Perspect. Psychol. Sci.* 15, 1011–1025
9. Rakow, T. (2022) Adversarial collaboration. In *Avoiding Questionable Research Practices in Applied Psychology* (O'Donohue, W. *et al.*, eds), pp. 359–377, Springer International Publishing

10. Ceci, S.J. *et al.* (2023) Exploring gender bias in six key domains of academic science: an adversarial collaboration. *Psychol. Sci. Pub. Interes.* 24, 15–73
11. Melloni, L. *et al.* (2021) Making the hard problem of consciousness easier. *Science* 372, 911–912
12. Bacon, F. (1878) *Novum Organum*, Clarendon press
13. Lakatos, I. (1974) The role of crucial experiments in science. *Stud. Hist. Phil. Sci.* 4, 309–325
14. Corcoran, A.W. *et al.* (2023) Accelerating scientific progress through Bayesian adversarial collaboration. *Neuron* 111, 3505–3516
15. Blohm, G. *et al.* (2024) Generative Adversarial Collavotations: a practical guide for conference organizers and participating scientists. *arXiv*, Published online February 19, 2024. https://doi.org/10.48550/arXiv/2402.12604