AMERICAN PSYCHOLOGICAL ASSOCIATION SARMAC Journal of Applied Research in Memory and Cognition

ISSN: 2211-3681

2022, Vol. 11, No. 1, 23–26
https://doi.org/10.1037/mac0000002

COMMENTARY

Adversarial Collaborations in Behavioral Science: Benefits and Boundary Conditions

Madalina Vlasceanu[1], Diego A. Reinero[2], and Jay J. Van Bavel[1]
[1] Department of Psychology, New York University, United States
[2] Department of Psychology, Princeton University, United States

Keywords: adversarial collaboration, behavioral science, boundary conditions

> (T)he antagonists must have a strong scientific curiosity and an honest desire to discover the truth, rather than being concerned primarily with protecting their pet theory against attack. Their self-esteem must be based on using the correct process to discover knowledge, rather than on getting the desired outcome (e.g., being right). They must be willing to look at the facts objectively—Latham et al. (1988, p. 771).

In 1988, Latham and colleagues presented "a method of resolving scientific disputes that may be unique in the history of psychology" in which scholars with a competing theoretical perspective worked together to design a crucial test of these perspectives with a third-party serving as an unbiased mediator (see Latham et al., 1988). Although this technique failed to catch on, it nevertheless provides a potentially fruitful approach to address contentious scientific issues (see Kahneman, 2003, 2011). In a recent article, Clark et al. (2022) emphasize the value of this research practice as a way to expedite science and truth-seeking, by having scholars who disagree with each other's interpretation of data or theory work together to resolve their dispute.

We agree that adversarial collaborations offer a potentially useful tool for scientists and should be used more often. Our experience conducting adversarial collaborations has given us some insight into potential challenges of this approach. In this commentary, we describe the benefits and boundary conditions of adversarial collaborations and propose suggestions for incentivizing border participation in this research practice. We hope our article will provide a useful roadmap for scholars who are considering implementing an adversarial collaboration, by offering guidance on avoiding some of the drawbacks.

Human Biases, Science's Limits, and the Benefits of Adversarial Collaborations

People's beliefs are shaped by both accuracy goals and social goals, and these two motivational sources can come into direct conflict (Van Bavel & Pereira, 2018; Van Bavel et al., 2020). When the stakes for believing accurate information are low, people often prioritize social goals such as conforming to a group's beliefs in order to fit in or rise in social status (Clark et al., 2022). In fact, even when the stakes for believing accurate information are high (e.g., getting vaccinated to reduce risk of serious illness or death during the coronavirus disease [COVID-19] pandemic), people sometimes still prioritize partisan beliefs (e.g., many Republicans refuse to get vaccinated or deny its efficacy), or even promote conspiracy theories (Douglas, 2021). Therefore, people are not always able to arrive at optimal choices.

Scientists are not exempt from these inherently human biases. Scientists might, for example, weigh the idealized pursuit of truth (an accuracy goal) against their own real-life constraints such as career concerns or social status (a social goal). Importantly, however, scientists do value accuracy and rigor (Kahan et al., 2017), require greater empirical consistency than nonscientists (Hogan & Maglienti, 2001), and very few engage in blatantly unethical behaviors (Stricker & Günther, 2019). More importantly, scientists embrace norms and institutions designed to mitigate against biased thinking (Merton, 1942; Van Bavel et al., 2020). These cognitive styles and norms might help mitigate against bias and allow scientists to generate more accurate beliefs than alternative belief systems.

Nevertheless, behavioral science may pose additional challenges to scientists. This type of work can be particularly prone to inconsistencies given that psychological constructs are hard to measure and can be measured in different ways (Landy et al., 2020), findings are subject to shifting contexts and populations (Van Bavel et al., 2016), and research topics carry political implications which can give rise to implicit ideological biases. Indeed, while the *pursuit* of

Madalina Vlasceanu https://orcid.org/0000-0003-2138-1968
Diego A. Reinero https://orcid.org/0000-0002-6124-2623

The authors declare no conflicts of interest.

Madalina Vlasceanu, Diego A. Reinero, and Jay J. Van Bavel contributed equally. Authorship order was determined by bets placed on the outcome of a National Hockey League (NHL) game (Edmonton Oilers at New York Rangers, March 1, 2022).

No data were collected for this article.

Correspondence concerning this article should be addressed to Madalina Vlasceanu, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, United States. Email: mov209@nyu.edu

value-free science is a laudable goal, its *full attainment* may be philosophically unachievable (Longino, 1990; Richardson & Polyakova, 2012; Rykiel, 2001; Sears, 1994). For this reason, behavioral scientists might benefit from additional institutional practices to help mitigate against these challenges.

Additionally, peer review, while critical in the advancement of science, is not perfect. And open science may not necessarily address the issue of resolving contradictory theories as it leaves open the possibility for researchers to define research questions and operationalizations of their choosing which can bias results and further augment scientific disputes. Likewise, current scientific reward structures can benefit scholars who only work with like-minded others and who, instead of directly working with adversaries, benefit from a sequential article-commentary debate.

Given the imperfections of scientists and the scientific process, adversarial collaborations can be a helpful strategy to resolve academic disagreements (though we do not believe they are the only effective strategy). Through concurrent dialogue among scientific adversaries, ideas, predictions, designs, and measures should, in theory, be established more clearly. This practice can also ensure a higher level of scrutiny since rivals are expected to review and agree to each step in the research process, which can guard against post hoc rationalizations or hindsight bias.

Finally, science may move forward faster and more efficiently when competing theories are put to a compelling test and the data reveal which one most closely resembles reality. When scientific rivals or adversaries engage with each other's ideas directly, rather than through rebuttals, science advances more quickly. In fact, some of the more popular and productive talks at scientific conferences are ones in which scientific rivals debate each other by laying out their strongest arguments, allowing the audience to draw conclusions about which data or theory seem most convincing. Such debates create a space in which scientists directly address each other, clarify points of confusion or misinterpretations, attempt to resolve conflicting evidence, and persuade the audience. Adversarial collaborations take this practice a step further, forcing disagreeing scientists to work together in carrying out a study (or set of studies) aimed at adjudicating the truth.

Boundary Conditions, Limitations, and Suggestions for Adversarial Collaborations

Despite the potential benefit of adversarial collaborations to the scientific process, the boundary conditions and limitations of this approach should be seriously considered. Here we discuss the range of utility, resource intensity, post hoc explanations, converging evidence, and need for peer review. These points should not discourage scientists from participating in adversarial collaborations, rather they provide additional context and considerations, as well as suggestions for incentivizing participation.

In our view, the future of adversarial collaboration requires that we address these concerns in terms of the norms, institutions, and incentive structures we build as a field. In the sections below, we outline some of these limitations and then propose concrete suggestions for overcoming these issues. With the right set of changes, perhaps the field will embrace this scientific approach and, in turn, increase the accessibility and scientific utility of adversarial collaborations as a research strategy.

Resource Intensity

One of the main barriers to adversarial collaborations is the higher level of resources they require compared to an average research project (Clark et al., 2022). The increase in time investment, effort, and uncertainty (as well as the stress from social conflict) when conducting a study with adversarial collaborators might discourage scientists, especially those in early career stages, from embarking on such a project. And even if scientists do commit to an adversarial collaboration, the lower degree of trust in an adversarial collaborator can lead to conflict and eventual dropout from the project. This is a considerable opportunity cost of participating in an adversarial collaboration, when compared to collaborating with trusted colleagues who share a common theoretical and methodological framework as well as existing relationships.

We believe this is one of the core barriers to conducting an adversarial collaboration and should be baked into the heart of any discussions about how to scale this activity as a core part of psychological science. This is a case where funding agencies and foundations should allocate more resources to resolving the central theoretical debates in a field, rather than funding singular research programs by a single lab (or research team). It is also the case that dissertation committees, hiring committees, and tenure and promotion committees need to appreciate the effort and value the intellectual contribution that these sorts of activities entail for the field. And scientific journals should allocate special issues or sections to adversarial collaborations. Until these incentive structures are in place, it seems likely that people will continue to opt for the easier, less expensive, and more enjoyable pathway of conducting more efficient confirmatory studies on their own pet theories.

Peer Review

Collaborating with a scientist with opposing hypotheses on a research project is not likely to render the scientific peer-review system obsolete or provide a path to circumvent it. Despite the adversarial nature of the project, given that all researchers involved will enjoy the professional benefits of authorship from a published research output, they are still subject to the biases the peer-review system was designed to deter (e.g., misrepresenting or misinterpreting the data, using inappropriate statistical analyses, overgeneralizing the findings, etc.). Anonymous third parties assessing the project from an impartial standpoint are still needed to ensure the scientific rigor of the research product.

One form of peer review that seems to be particularly useful for adversarial collaborations would be preregistered reports. In this format, the peer reviewers have input into the design of the project before data collection. This can complement the discussions conducted by the adversarial team, address any further blind spots among the team, and perhaps help resolve any disputes between rival collaborators. More importantly, this commits the editor and journal to evaluate and accept the article based purely on the importance of the question and methodological design features. As such, it provides a framework for accepting and publishing the final article even if the results are messy or incongruent with the perspective of any of the adversarial collaborators. It also provides a clear record of decision-making as part of the review process. Finally, this might also be useful if the issue is contentious, and the editor has either a stake in the outcome or strong priors about the outcome.

Range of Utility

A core motivation for adopting adversarial collaborations is the avoidance of scientific misconduct or questionable practices researchers might engage in when pressured by social motives such as career aspirations (Clark et al., 2022). Although such instances do exist, their prevalence is often overestimated. For example, the prevalence of questionable research practices in psychology had been overestimated by several orders of magnitudes, a result of using ambiguous and misleading response formats to investigate such practices (Fiedler & Schwarz, 2016). Moreover, recent empirical investigations revealed that only 0.0082% of journal articles in psychology were retracted due to scientific misconduct (Stricker & Günther, 2019). Relatedly, an empirical assessment of the file drawer problem, which assumes that null results are less likely to be included in publications and thus in meta-analyses, concluded that the file drawer problem does not actually produce significant biases in estimating effects (Dalton et al., 2012). Therefore, although useful in deterring questionable practices in psychology, the low frequency of such practices narrows the range of utility of adversarial collaborations to boundary cases.

Given the potential costs of conducting adversarial collaborations, we should give greater consideration to their range of utility. To this end, the gatekeepers for funding adversarial collaborations (as well as individual researchers) should consider a number of questions before determining if this direction is worth pursuing. Instead of focusing on addressing fraud, we believe they should instead focus on topics that are likely to benefit from an adversarial collaboration. First, the issue should have significant theoretical or practical implications for the field and beyond. Second, the topic should be the center of spirited debate between multiple researchers or lab groups—where it is hard for an outsider to determine the consensus view. Third, it should be a topic where it could benefit from preregistration (which is an essential part of the adversarial process). And fourth, it should be a topic where members of the collaboration could be convinced to update their beliefs pending the results of the project. For instance, recent research shows that people update their beliefs most when they are confronted with large prediction errors, even across ideological boundaries (Vlasceanu et al., 2021). Accordingly, an optimal adversarial collaboration topic is one that prompts predictions as different from each other as possible. When these conditions apply, the topic would likely be worthy of the additional effort and resources necessary for an adversarial collaboration.

The Value of Converging Evidence

Another limitation of adversarial collaborations in settling scientific disputes is the dependence on a unique operationalization of the construct of interest, and a unique method of assessing the outcome of the debate. Converging evidence relying on distinct operationalizations and multiple methods generally offers the most conclusive evidence in the scientific quest for truth (Hedges, 2000). For instance, experimental research methods which have high-internal validity but low external validity can be complemented by field studies high in external validity but low in internal validity. While coming to a consensus among the adversarial collaborators regarding a single operationalization is the traditional goal of an adversarial collaboration, it might not prove compelling to other experts in that specific community.

There is scientific value in the multiple operationalizations, multiple approaches, and converging evidence across multiple levels of analysis (Wilson, 1999), and adversarial collaborations will need to balance this tenet of science against the need for a single test aimed at solving the debate. We encourage adversaries to design the project in a way that will motivate belief updating among the broadest possible audience. One way to do this is by using multiple operationalizations of the key theoretical variables. This will allow multiple simultaneous tests of the core theoretical question, without allowing opponents to dismiss the results as a function of a single operationalization. This would also make the findings more generalizable and theoretically incisive, by addressing the broader latent constructs at the core of the theoretical debate. Not only will this type of approach be more convincing to adversarial collaborations, but it is also more likely to provide a compelling body of evidence for third parties who are committed to a preexisting perspective. We believe this external audience might be especially important to consider since the conversation between adversaries will likely tend to focus on their own perspectives.

Post Hoc Explanations

In theory, adversarial collaborations offer a solution to debates between researchers with opposing views about specific issues (Clark et al., 2022). The logic is that if both sides agree to a method of testing the contentious hypothesis, the data will settle the debate. In reality, this ideal scenario is almost perfectly designed to be threatening, which can motivate people to dismiss or explain away evidence that contradicts valued beliefs (Sherman & Cohen, 2006; Steele, 1988), that increases cognitive dissonance (Festinger & Carlsmith, 1959), and that reduces coherence among already held beliefs (Lord et al., 1979). Therefore, when the stakes are high, it is possible researchers will explain away data that invalidate their theories by questioning the methods and analyses they initially signed off on. We propose that one way of addressing this concern is to encourage the preregistration of each collaborator's predictions of the results based on the design agreed upon. This point prediction would enforce additional accountability by quantifying collaborators' expectations before the data have been collected.

However, we suspect this is a larger problem for third-party observers with committed beliefs than actual adversarial collaborators. In our experience, this is likely to be one of the strongest barriers to the utility of a collaboration. In a recent adversarial collaboration on the role of political bias in replication (Reinero et al., 2020), although all members of the adversarial collaboration team were convinced by the utility of the data, the findings were criticized by reviewers and online readers with strong priors against the conclusion. This example suggests that the challenges of convincing firmly committed third parties might be intractable through this framework. Therefore, we believe the focus should be on convincing the adversarial collaborators and unbiased third parties who are open minded to the results and attuned to the rigor of the experimental process rather than to a particular outcome.

Conclusion

We firmly believe that adversarial collaborations have the *potential* to help resolve contentious scientific debates. To achieve this potential, however, we think that scientists considering this approach should understand and confront the potential barriers and challenges to successful adversarial collaborations. Until these issues are addressed, this approach will likely remain too expensive, difficult, and unpersuasive to compel researchers to participate. As such, it will likely remain a rare, niche exercise among scholars who are willing to address controversial questions. However, incentivizing this practice using our suggestions has the potential to make it more normative and appealing for a wider variety of scientists.

References

Clark, C., Costello, T., Mitchell, G., Tetlock, P. (2022). Keep your enemies close: Adversarial collaborations will improve behavioral science. *Journal of Applied Research in Memory and Cognition*, *11*(1), 1–18. https://doi.org/10.1037/mac0000004

Dalton, D. R., Aguinis, H., Dalton, C. M., Bosco, F. A., & Pierce, C. A. (2012). Revisiting the file drawer problem in meta-analysis: An assessment of published and nonpublished correlation matrices. *Personnel Psychology*, *65*(2), 221–249. https://doi.org/10.1111/j.1744-6570.2012.01243.x

Douglas, K. M. (2021). COVID-19 conspiracy theories. *Group Processes & Intergroup Relations*, *24*(2), 270–275. https://doi.org/10.1177/1368430220982068

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, *58*(2), 203–210. https://doi.org/10.1037/h0041593

Fiedler, K., & Schwarz, N. (2016). Questionable research practices revisited. *Social Psychological & Personality Science*, *7*(1), 45–52. https://doi.org/10.1177/1948550615612150

Hedges, L. V. (2000). Using converging evidence in policy formation: The case of class size research. *Evaluation & Research in Education*, *14*(3–4), 193–205.

Hogan, K., & Maglienti, M. (2001). Comparing the epistemological underpinnings of students' and scientists' reasoning about conclusions. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *38*(6), 663–687. https://doi.org/10.1002/tea.1025

Kahan, D. M., Landrum, A., Carpenter, K., Helft, L., & Hall Jamieson, K. (2017). Science curiosity and political information processing. *Political Psychology*, *38*(Suppl. 1), 179–199. https://doi.org/10.1111/pops.12396

Kahneman, D. (2003). Experiences of collaborative research. *American Psychologist*, *58*(9), 723–730. https://doi.org/10.1037/0003-066X.58.9.723

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus, and Giroux.

Landy, J. F., Jia, M. L., Ding, I. L., Viganola, D., Tierney, W., Dreber, A., Johannesson, M., Pfeiffer, T., Ebersole, C. R., Gronau, Q. F., Ly, A., van den Bergh, D., Marsman, M., Derks, K., Wagenmakers, E. J., Proctor, A., Bartels, D. M., Bauman, C. W., Brady, W. J., … the The Crowdsourcing Hypothesis Tests Collaboration. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. *Psychological Bulletin*, *146*(5), 451–479. https://doi.org/10.1037/bul0000220

Latham, G. P., Erez, M., & Locke, E. A. (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. *Journal of Applied Psychology*, *73*(4), 753–772. https://doi.org/10.1037/0021-9010.73.4.753

Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton University Press. https://doi.org/10.1515/9780691209753

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098–2109. https://doi.org/10.1037/0022-3514.37.11.2098

Merton, R. K. (1942). The ethos of science. *Journal of Legal and Political Sociology*, *1*, 115–126. (Reprinted from *Social structure and science*, by R. K. Merton and P. Sztomka, Ed., 1996, University of Chicago Press).

Reinero, D. A., Wills, J. A., Brady, W. J., Mende-Siedlecki, P., Crawford, J. T., & Van Bavel, J. J. (2020). Is the political slant of psychology research related to scientific replicability? *Perspectives on Psychological Science*, *15*(6), 1310–1328.

Richardson, E. T., & Polyakova, A. (2012). The illusion of scientific objectivity and the death of the investigator. *European Journal of Clinical Investigation*, *42*(2), 213–215. https://doi.org/10.1111/j.1365-2362.2011.02569.x

Rykiel, E. J. (2001). Scientific objectivity, value systems, and policymaking. *Bioscience*, *51*(6), 433–436. https://doi.org/10.1641/0006-3568(2001)051[0433:SOVSAP]2.0.CO;2

Sears, D. O. (1994). Ideological bias in political psychology: The view from scientific hell. *Political Psychology*, *15*(3), 547–556. https://doi.org/10.2307/3791572

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 38, pp. 183–242). Academic Press.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 261–302). Academic Press. https://doi.org/10.1016/S0065-2601(08)60229-4

Stricker, J., & Günther, A. (2019). Scientific misconduct in psychology: A systematic review of prevalence estimates and new empirical data. *Zeitschrift für Psychologie*, *227*(1), 53–63. https://doi.org/10.1027/2151-2604/a000356

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(23), 6454–6459. https://doi.org/10.1073/pnas.1521897113

Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in Cognitive Sciences*, *22*(3), 213–224. https://doi.org/10.1016/j.tics.2018.01.004

Van Bavel, J. J., Reinero, D. A., Harris, E., Robertson, C. E., & Pärnamets, P. (2020). Breaking groupthink: Why scientific identity and norms mitigate ideological epistemology. *Psychological Inquiry*, *31*(1), 66–72. https://doi.org/10.1080/1047840X.2020.1722599

Vlasceanu, M., Morais, M. J., & Coman, A. (2021). The effect of prediction error on belief update across the political spectrum. *Psychological Science*, *32*(6), 916–933. https://doi.org/10.1177/0956797621995208

Wilson, E. O. (1999). *Consilience: The unity of knowledge* (Vol. 31). Vintage.

Received January 11, 2022
Accepted January 12, 2022 ■