PERSPECTIVE

Adversarial alignment enables competing models to engage in cooperative theory building toward cumulative science

Naomi Ellemers[a], Susan T. Fiske[b,1], Andrea E. Abele[c], Alex Koch[d], and Vincent Yzerbyt[e]

Edited by Jennifer A. Richeson, Yale University, New Haven, CT, and approved February 18, 2020 (received for review September 2, 2019)

Crises in science concern not only methods, statistics, and results but also, theory development. Beyond the indispensable refinement of tools and procedures, resolving crises would also benefit from a deeper understanding of the concepts and processes guiding research. Usually, theories compete, and some lose, incentivizing destruction of seemingly opposing views. This does not necessarily contribute to accumulating insights, and it may incur collateral damage (e.g., impairing cognitive processes and collegial relations). To develop a more constructive model, we built on adversarial collaboration, which integrates incompatible results into agreed-on new empirical research to test competing hypotheses [D. Kahneman, *Am. Psychol.* 58, 723–730 (2003)]. Applying theory and evidence from the behavioral sciences, we address the group dynamic complexities of adversarial interactions between scientists. We illustrate the added value of considering these in an "adversarial alignment" that addressed competing conceptual frameworks from five different theories of social evaluation. Negotiating a joint framework required two preconditions and several guidelines. First, we reframed our interactions from competitive rivalry to cooperative pursuit of a joint goal, and second, we assumed scientific competence and good intentions, enabling cooperation toward that goal. Then, we applied five rules for successful multiparty negotiations: 1) leveling the playing field, 2) capitalizing on curiosity, 3) producing measurable progress, 4) working toward mutual gain, and 5) being aware of the downside alternative. Together, these guidelines can encourage others to create conditions that allow for theoretical alignments and develop cumulative science.

theory building | scientific competition | cooperation | negotiation | behavioral science

Leading representatives of five rival scientific models gathered to identify common views, explain some contradictions, and acknowledge lasting disagreements. Aiming to find a constructive alignment process, we applied insights from the behavioral sciences to specify boundary conditions and communication rules, creating a situation that would facilitate adversarial alignment. This was possible by defining the joint goal of cooperating toward theoretical progress, by respecting each other's scholarly competence, and by trusting each other's good intentions in pursuing that goal. Techniques borrowed from negotiation research provided a conducive context for scientific debate focusing on content. We believe that this paradigm might be useful to others facing conceptual conflict.

The Challenge

To illustrate the challenge, consider five competing theories that make contradictory predictions about the fundamental dimensions of social evaluation. As social creatures, people constantly evaluate themselves, other individuals, and groups. This is arguably functional for guiding behavior, even for surviving and thriving in the social world. The theories differ on the number, meaning, priority, relation, function, and processes of these evaluative dimensions (see Box 1). A body of published evidence substantiates each theory across domains. After decades of research, however, no single model dominates. The results of this collaboration—content of arguments, nature of observations, theoretical convergence, remaining disagreement, and novel

[a]Faculty of Social Sciences, Utrecht University, Utrecht, 3508 TC, The Netherlands; [b]Psychology and Public Affairs, Princeton University, Princeton, NJ 08544; [c]Social Psychology, University of Erlangen-Nümberg, Bayreuth D91054, Germany; [d]Booth School of Business, University of Chicago, Chicago, IL 60637; and [e]Institute for Psychological Sciences, University of Louvain, B-1348 Louvain-la-Neuve, Belgium

Author contributions: N.E., S.T.F., A.E.A., A.K., and V.Y. provided analytic framework and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: sfiske@princeton.edu.

Box 1. Controversies and Alignment of Five Competing Models for Social Evaluation

Social evaluation—judging self and others—is inescapable, ubiquitous, and crucial for navigating social interaction. The research has a long history in social psychology. Comparing five current models for social evaluation revealed three key controversies. The models and their supporting evidence disagree about 1) the social evaluative dimensions' number, organization, and definition/labels; 2) the possible priority of one dimension; and 3) the statistical relation among the dimensions.

Specifically, the Dual Perspectives Model (51, 52) primarily addresses evaluations of self vs. others, maintaining that "agency" is primary in evaluations of the self and that "communion" is primary in evaluations of the other. The Behavioral Regulation Model (53, 54) focuses on the self in relation to ingroup and outgroup members and finds that "morality" weighs more heavily in overall judgments than do "competence" or "sociability." The Dimensional Compensation Model (55, 56) addresses situations comparing two targets (usually groups) and finds that differences on competence are often compensated by differences on "warmth" (and vice versa). The Stereotype Content Model (24, 57) compares stereotypes of multiple groups in society by locating them in a two-dimensional warmth by competence space. The Agency, Beliefs, and Communion Model (28, 58) captures perceptions of similarity in agency and (conservative-progressive) "beliefs" between the self and many different societal groups, which elicit inferences about the communion of these groups.

We made an effort to accept published research evidence while displaying curiosity and gathering information about each model's theoretical roots, focal domains, operationalizations, and research methods. This effort was crucial to explain why the nature, focus, and premises of the models differ. As a first step in our discussion toward alignment, we agreed that the models can to some extent be seen as complementary because of speaking to different contexts and targets, ranging from single individuals to large numbers of groups in society.

Closer consideration of the conceptualizations and evidence pertaining to Controversy 1 (number, organization, and definition of dimensions) led to our specifying two overarching dimensions (horizontal and vertical), each containing different facets (morality and friendliness as well as assertiveness and ability, respectively). This allowed partial alignment of the five models and yielded additional predictions (e.g., specifying when the distinction between facets would be important) to be tested in future research.

Consideration of Controversy 2 (dimensional priority) and how this was assessed by the different models led to the distinction between priority indicators assessing processing speed, subjective weight (both prioritize horizontal), and pragmatic diagnosticity (prioritizes vertical) and to specifying testable predictions about moderators (actor vs. observer perspective, personal interaction, degree of interdependence in an interaction, analytical insight goals) that would affect priority.

Aligning the models to address Controversy 3 (relation between dimensions) was possible by considering different relations (orthogonal, positive, negative, or curvilinear) generated by the number of targets considered; alignment also required distinguishing between the semantic meaning of the dimensions and their valence on a positivity–negativity continuum.

Altogether, a number of resulting insights illuminated the processes of social evaluation (catalysts, antecedents, consequences) as well as the functions (e.g., social comparison, judgment accuracy, self and ingroup esteem, and guiding interaction). The integrative alignments—and remaining disagreements—are generating research as well as having practical significance (59).

hypotheses—are reported elsewhere (*Controversies and Alignment of Five Competing Models for Social Evaluation* has an overview). The current contribution specifies key aspects of the social interaction that need to be managed to allow for a discussion about areas of disagreement, resolve controversies, and reach clarity about remaining points of contention. Below, we draw on theory and research from the behavioral sciences as we report the process and specify some preconditions and rules of engagement that enabled it.

Never Waste a Good Crisis: Reframing Adversaries as Collaborators

The project developed in a larger climate of scientific crisis about transparency and reproducibility (1). Granted, dissent and debate are tools that can advance thinking and findings in all disciplines of science. Our models' disagreements were substantive, involving both evidence and theory. Competing viewpoints can lead to the design of new experiments that allow theoretical progress; this is the philosophy of science that orients many researchers. However, current debates also relate to the ground rules of doing science (1). Some issues are general (appropriate research samples, methods, statistics, transparency, accessibility), and some are domain specific (comparability of results, ethical implications of research procedures). In this context, different viewpoints and seemingly incompatible findings easily elicit skepticism (were the other researchers biased?), suspicion (were their procedures appropriate?), or confusion (how seriously should we take these findings?).

Science progresses not only by challenging the generality, robustness, and consistency of empirical results. Science also advances from efforts to understand how seemingly incompatible findings can be resolved, how more refined techniques allow deeper insight into mechanisms underlying classic findings, and how variations in research approaches, designs, or measures may uncover moderating factors that help establish the generality vs. specificity of prior observations (2). Yet, analyses of publication trends reveal that a focus on optimizing empirical techniques can have the side effect of reducing the range of theoretical issues under examination (3).

Thus, if science has been in crisis, this crisis concerns not only methods and statistics but also, theory development (4, 5). Beyond refining tools and procedures, reconciling contradictory results requires a deeper understanding of the theoretical underpinnings guiding research. In practice, however, theories rarely collaborate; using other people's theory has been equated to using their toothbrush (6). Because people own them, theories compete, and requiring some to lose for others to win incentivizes destruction of—rather than curiosity about—seemingly incompatible views. Just publishing contradictory viewpoints and incompatible data does not, however, advance science. Progress requires connecting dots, lumping evidence, and bridging perspectives (7–15).

Progress toward Cooperative Joint Goals: Starting with Trust

Science is often compared with competitive sports. The competitive scientific climate is reinforced by scarcity of positions and resources as well as shared beliefs about the best way of doing science—supported by common references to sports metaphors, such as the importance of "winning" or "getting there first." However, such climates also have known disadvantages. Research on motivation and achievement at work shows that competitive incentive systems easily elicit envy and mistrust and are demonstrable sources of inefficiency, obstruction, and other undesirable work behavior (16–19). This is not different in science (20, 21). In countries that rely more on competition between researchers to allocate resources for research, the relation between input (funding awarded) and output (highly cited research publications) is less favorable than in less competitive scientific systems (22). For opposing teams, competition can impair information processing (23). Additionally, unfettered public competition can undermine collegial relations; competition creates mistrust (24).

The authors of this joint essay have each made public claims regarding their own approach, critiquing the others. In the context of such public disagreements, negotiating consensus requires revisiting public commitments. Notwithstanding the shared conviction that progress in science requires rejecting and replacing prior understandings of reality, the reconsideration of publicly expressed views is notoriously challenging—even for scientists. Couching scientific viewpoints as competing models championed by opposing teams intertwines logic and evidence with scientific identities and reputations. This mix discourages trusting others' motives and being open to their critique (25). Yet, advancement through debate is only possible when people can avoid being defensive and can reconsider their own position when offered a new perspective (26). The sports metaphor does not take into account that in science—unlike sports—excellence is not established on a single unambiguous criterion (27). Diversity of perspectives is key to innovation and creativity (28) that characterizes progress in science.

We propose that cooperation between divergent viewpoints can help move away from a zero-sum representation of progress in science. Even if areas of disagreement remain, cooperation can expand the pie by introducing additional dimensions, considerations, and perspectives to create novel research predictions that can develop cumulative theory. Collaborating with others who contribute diverging insights and bodies of evidence can achieve this, but it is notoriously difficult and requires trust in their intentions. Mutual trust generally predicts successful cooperation, innovation, and team performance, while lack of trust prevents benefiting from each other's professional skills (29, 30).

Scientists' Rules of Engagement: Steps to Successful Negotiation of Theory

Currently, no clear guidelines describe how research groups can manage competing theories. As a constructive response to contradictory evidence and failed replications, adversarial collaboration on research has integrated incompatible results (31). Inspired by this idea, we proposed to build an "adversarial alignment" of competing conceptual frameworks. Any resulting metatheoretical consensus—as well as clarity about remaining disagreements—would cumulatively advance science. Insights from theory and research on the psychology of negotiation and conflict resolution address the group dynamic complexities of managing such adversarial interactions. Below, we identify five key conditions that allow for collaboration between adversaries and illustrate how we secured these in our case (32, 33).

Leveling the Playing Field. Differential power, status, and seniority influence scientists' perceptions, emotions, and commitment to their own and each other's ideas. This can discourage a fair exchange of information, resources, and consideration of everyone's contributions (34). Indeed, the number one recommendation for a successful negotiation outcome is to "separate the people from the problem." To achieve this, we assumed that each approach was valid and deserved inclusion based on each having published evidence from multiple studies. Furthermore, we decided to have one individual act as a representative for each approach, with the result that five rivals would meet face to face in a neutral nonacademic location (in our case, a hotel)—and share the outcome of the alignment with their collaborators only after this initial meeting. The individual-to-individual approach facilitates flexibility.

Our group included one early-career, two midcareer, and two more senior people. Eliminating differences in academic experience, status, and power is challenging. Different career stages elicit different concerns, and at each stage, scientists need others to recognize the validity of their ideas and to acknowledge their contributions. For junior as well as more senior colleagues, making oneself vulnerable to the critique of colleagues took nerve. Yet, it was the only way forward, even if it made us all feel uncomfortable. Sharing this experience only with each other outside the scrutiny of our coworkers and without a broader audience (as at a typical scientific meeting or conference) released us from these audience concerns and allowed for a collaboration on equal footing.

One of the first decisions we made to secure a voice for each of us was that we would prepare a joint paper for which authorship would be alphabetical. We ensured that all authors would be allowed to contribute equally by monitoring equal space in the paper to be devoted to each approach. The person representing that approach would be the lead author on his or her own section, with editorial comments from others serving only to clarify the writing.

Capitalizing on Curiosity. Studies on conflict management reveal that people are well able to handle disagreement about specific outcomes (in our case: divergent findings rooted in different theoretical premises) as long as they share the same underlying values (in our case: seeing merit in tailoring empirical approaches to examining specific questions) (35). A second key agreement, which we made ahead of time, therefore was to accept the peer-reviewed empirical results reported by each of the research groups as valid contributions to the literature and to restrict the debate to issues of generalizability and interpretation. This aligns with the second rule for successful negotiation, recommending that parties "focus on interests, not positions." Instead of challenging each party to defend what they had found and concluded from their findings, we invited each other to specify both (at a higher level of abstraction) why they had chosen to examine particular phenomena and (at a lower level of abstraction) how they were trying to understand them with their particular research samples, measures, and procedures. This curiosity about the origins and nature of our previous inquiries at different levels of abstraction greatly helped with developing our adversarial cooperation.

We operationalized this part of the collaborative effort by each preparing a summary statement ahead of time detailing the

theoretical and empirical origins of our respective approach and how we developed our thinking from there. We started our face-to-face time by aligning in tables the what, why, and how of our five research approaches.

Producing Progress. Research reveals that team collaboration benefits from dedication to joint tasks. This is facilitated by practical arrangements that require costly commitment and make parties agree on a shared goal ahead of time (36). Furthermore, studies on successful goal pursuit document that breaking down abstract objectives into smaller concrete tasks enhances feelings of self-efficacy that support motivation and facilitate goal achievement (37). Explicitly assigning tasks and roles that acknowledge complementary skills and contributions (knowledge of specific literatures, expertise in particular methods and statistical techniques, publication experiences, access to resources for consultation and assistance, writing skills) make it possible to benefit from differences instead of suffering from them (38). Face-to-face interaction helps to avoid miscommunication and build trust, which facilitates subsequent remote meetings and long-distance collaboration over time (29). Considering "devil's advocacy" as a contrarian technique is well known in negotiation research and generally serves to prevent premature solutions (39).

We created commitment to the collaboration and successfully advanced toward our shared goal by defining concrete and practical steps on how to progress toward a joint outcome using a different strategy in each phase of our collaboration.

Preparation. The preparation phase started about a year in advance by setting aside five workdays to get together in an isolated hotel, with the goal to complete a first draft of a joint theory paper by the end of that week. A month before the meeting, each prepared a reading list about his or her perspective as well as a five-page homework answering several standard questions (origins, insights, premises, methods, evidence). This drafted a third of the paper and made the task seem more feasible. The written preparation forced each model to use the same, compatible format to provide a concise overview of its main theoretical statements, to summarize empirical data dispersed across different outlets, and to specify the choice of research designs and measures. Even before the meeting started, each of us had already invested time in preparing and writing, and each had committed to work together for a number of days without distraction.

Meeting. At the meeting, to be able to make progress, we were careful to delay talking about disagreements. The homework for the meeting offered a unified format that allowed us to compare what each of our theories—apart from studying social evaluation—dealt with more specifically and how we had examined the theory. During the first days, we briefed each other and named areas of agreement. At this stage, we did not devote a lot of time to issues that we could not reconcile; these we relegated to a separate list to confront later. We did not just talk but during the discussion, jointly planned a paper with different parts, drafted tables and graphs, and prepared a set of slides (projected on the beamer while talking) outlining key arguments, definitions, points of consensus, and likely moderators. Only after committing to this common ground and having secured progress did we then proceed: documenting different domains of interest, specifying distinctive methods, and developing new predictions offering guidance to researchers wishing to further examine our (dis-) agreements.

First draft. During the second half of the meeting, dyads drafted sections of the paper in breakout sessions and circulated these to feed follow-up conversation. We delayed searching for and discussing nonnegotiable disagreements until the end of our meeting. The week was not long enough to craft remaining points of controversy or to specify how future research might further test them. However, the effort invested in creating a cooperative context, the confidence developed in each other's complementary contributions, and the optimism built in our ability to negotiate points of convergence enabled us to continue after we separated. This allowed us to prepare a first full draft of a joint manuscript soon after the meeting.

Feedback and revision. After completing the raw version of our manuscript, we each consulted our respective collaborators. They generally supported our endeavors but also, encouraged us to sharpen the controversies to benefit more from the theoretical alignment. With this feedback, they in a certain sense served as devil's advocates who ensured that alignments were not premature and that unresolved conflicts would be clearly spelled out. After having collected this feedback, we met on Skype to coordinate multiple rounds of revision and worked together long distance in dyads to improve sections of text. After the meeting, we needed almost a year of on-and-off collaboration to achieve this. We followed the same strategy to revise the paper in response to editorial comments after submission, which took another few months.

Working toward Mutual Gain. Negotiations are more likely to be successful when they create outcomes that benefit all parties involved. Collaboration between representatives of different groups is generally facilitated when the contact experience is positive (40, 41) and allows parties to define a shared identity (42, 43).

At the task level, we invested in formulating shared goals and facilitating the achievement of mutually beneficial outcomes. In addition to the added value of aligning five different models, each of our approaches benefited from this exercise. The curious and cooperative attitude that we adopted led to a better articulation of the distinct scope and added value of each model in developing more general insights about our joint concern. Having people work in dyads on specific subproblems—writing assignments that explain shared viewpoints and taking turns clarifying each other's text—ensured that each had a chance to help forge and own the end result. The different strategies we used all involved taking the perspective of another party and facilitated the development of joint ownership of the end result (44). For instance, we required each of us to consider our respective model's main reasoning and findings from the perspective of another theoretical framework, with the aim of developing novel insights. To avoid the aversive metaphor of "using someone else's toothbrush" (6), we referred to this as "taking your pet theory for a walk." This not only conveys discovering new scenes in this way but also, implies that novel perspectives can then come home for further consideration. Each of us had legitimate concerns about others misunderstanding complex issues that had been so carefully developed and about reaching premature alignment on issues our collaborators might not endorse.

To be able to continue our adversarial alignment, despite differences of opinion and scientific controversy, required finding a fun activity that periodically provided a chance for everyone to relax together. In our case, conviviality achieved this: sharing dinner at a different restaurant after every working day. Bonding is a nontrivial contributor to effective collaboration. A celebratory dinner planned for the last day symbolized the beneficial outcome we all wanted to achieve. (Others might pursue this by making

music, going for hikes, or visiting a different museum at the end of each day.) Having fun together and devoting some time to play are not a waste of working time; they regulate positive mood and make the experience rewarding.

Awareness of the Alternative. Successful negotiation has to improve on the default outcome. In the end, the motivation and ability of conflicting parties to reach mutual agreement depends on whether they are able to find a solution that is superior to the BATNA (best alternative to a negotiated agreement) (45). In our case, our joint presence at a prior international symposium made us understand that the alternative was not attractive. At this previous event, other scientists, on the one hand, expressed an interest in the topic of social evaluation and wanted to use one or more of our individual perspectives as a theoretical foundation for their research. On the other hand, they reported being aware of and confused by theoretical disagreements and voiced concern about being criticized by representatives from competing approaches. Thus, we had to contend with a BATNA that researchers in the discipline might turn away from all of our perspectives for fear of getting things wrong or just general inconclusiveness unless we could reconcile our own differences and indicate how new theory and research ideas might continue to develop.

The distinctive opportunity we carved out with social isolation, remote location, tight schedule, and extensive preparation not only created a feeling of commitment but also, gave us a sense of urgency in finding a way to discuss and resolve our controversies. The joint awareness of the undesirable alternative also resulted in the agreement that the representative of each perspective would take responsibility for securing consent from his or her respective constituents. We each approached those with whom we typically worked together in the past on developing the ideas that we represented. We asked them to scrutinize the accuracy of the rendition of our prior perspective in particular and to engage their view of the theoretical alignment as devil's advocates (39). However, as indicated in the description of different phases under step 3, we invited them to comment only after completing the first full draft of the joint paper. This also reflects a well-tested strategy to ensure successful representative negotiation by isolating representatives from their constituents to find a mutually satisfactory solution before reporting back and getting approval for the negotiated agreement (46).

Assessment: Hits, Misses, Correct Rejections, and False Alarms

Admittedly, many paths could lead to productive adversarial collaboration. Its success is not measured by the degree of alignment but by the achievement of clarity about specific points of contention and agreement about how to further examine these. Addressing the five steps outlined above should benefit other adversaries' ability to collaborate toward further development of science in this way. To be sure, the translated combination of negotiation theory derivations that worked for us would need adaptation to fit other situations (e.g., a larger or smaller number of competing frameworks) or might not work under other preconditions (e.g., less initial trust or more fundamentally diverging viewpoints). Hostility between adversaries or lack of curiosity about alternative viewpoints can prevent them from consenting to such a collaborative effort, and other chance events of an adversarial alignment week could also interfere. To paint a more complete picture, we here assess what processes went right and wrong. In accord with signal detection theory (47), below we highlight not only the negotiation theory aspects that our week laudably realized (hits) and laudably avoided (correct rejection) but also, aspects that our week unfortunately realized (false alarms) and unfortunately avoided (misses).

Hits or successes in the current approach meant that—as a consequence of the complex five-party negotiation on what to say and how to say it—we all accepted limits in how much each of us could press our favored text. The requisite care and diplomacy were only possible against the background of mutual trust and respect we had created. We realized how important it was to attend to the process of collaboration as well as its outcome, and we have continued working according to these rules in finalizing the theory paper after the meeting. At each stage, we also specified the division of tasks and contracted in writing what should be done by whom and when. Another important aspect for reaching hits was that we took time to come to conclusions. Too much pressure toward aligning controversies might lead to premature solutions that do not hold in the long run (32). We organized our long-distance cooperation after the meeting within an agreed-on timeframe for finishing the paper but took as many feedback loops as necessary to come to a satisfying result. In our case, this almost took another year. The walking our pet theories metaphor proved useful to think through further implications of combining different viewpoints. The term pet theory captured that we each had special concern about our favored perspective without making this too contentious.

Specifying the home turf of our respective models reflects the emergent realization that each originated from a specific concern and approach that would not necessarily expand to other areas of inquiry, a key observation in and of itself. Just having to take a walk allowed us to focus on the most obvious implications first and alleviated the pressure of having to be complete in reforming or connecting all aspects of our theories.

Finally, we collaborated in the realization that each of our models did not rely on a single set of results but was supported by a larger body of evidence representing the efforts of multiple laboratories. Considering larger programs of research in this way allowed us to accept that published work reflects the state of the art at the time, with research methods and analytical tools becoming more sophisticated over the years.

Correct rejections included the decision not to broaden the group (e.g., by inviting multiple representatives of each theoretical perspective to the meeting). We think this two-step process (with representatives reporting back to their constituents only after the meeting) was a crucial factor in the success of our collaboration and its result. In line with what negotiation theory would have predicted, having each perspective represented by only one individual made us more flexible and efficient, and it prevented the formation of subgroups or factions. This decision was not self-evident and even costly.

When preparing for the meeting, we put together a proposed meeting program and work plan to explore funding options. In itself, this was a useful exercise that provided an initial focus and statement of our common goal and joint ambition. However, to qualify for external funding options, we either had to meet at a time or venue that was inconvenient or had to expand the meeting goals and number of participants. Deciding against this meant that we had to self-fund. Each of us could find some way to cover travel expenses from research funds or university resources. Meeting near the home of one of the participants and putting up four people for four nights in a hotel on a remote location were not overly expensive, and one of us was able to cover this from

funds available to attract international experts for a research visit. We did not incorporate additional perspectives—our task was sufficiently difficult at this stage. However, the collaborative model and procedure we developed might be used in an iterative cycle, repeating at increasing levels of abstraction and representation, in an attempt to lump together even larger bodies of theory and data.

False alarms mainly related to the concerns we had ahead of time, namely whether the task we had set ourselves might be achieved. Specifying the theoretical goals and empirical origins of the five models before we assembled made us realize that we had been focusing on different aspects of a common larger issue. Having this conversation—and realizing that we approached similar issues from a different perspective—quickly alleviated this feasibility concern and was a first step toward the resolution of our diverging viewpoints.

We also had some concern about sharing the initial result of our efforts with different groups of collaborators. We feared they might be offended (about not being involved), difficult (about their work not being adequately represented), or defensive (of their own preferred perspective). None of these things happened. Our collaborators understood the added value of what we tried to do as well as the limitations we encountered. The comments we received mainly revealed curiosity about the nature of our efforts, openness about the approach taken, and encouragement to follow up on the progress we had made.

Misses resulted from our initial focus on finding points of agreement and defining common ground. This made us go overboard in highlighting the convergence and compatibility between the different approaches. As a result, the extent of the theoretical controversy was not sufficiently clear in the first draft of our text, and this was pointed out by several of our collaborators as well as reviewers. Fortunately, this was relatively easy to fix later on after we realized what had happened. In fact, we are convinced that following the reverse procedure (focusing on differences first and then finding points of agreement—which is more consistent with the standard paradigm in science) would have made it much more difficult to achieve our aims.

The 5 d allowed for aligning our perspectives but were too short to achieve real synthesis in carving out new theory and predictions (specifying the effects of particular moderators) or designing additional studies that might test these. The task was big and compressed in time; new insights were fresh and required further deliberation. Only after the meeting did we specify key controversies and consider how these could be resolved. Only after completing a first full draft of our paper and sending it out for comments were we able to develop new predictions and consider additional questions in more detail. This brought us to a deeper level of analysis and understanding and yielded new research ideas. Here, the care we took to specify our rules for engagement were instrumental in allowing us to continue our collaboration after the meeting. After spending the week together and better understanding each other's viewpoints and arguments, we could achieve this outcome long distance and forge new research collaborations aiming to test predictions emerging from our alignment efforts.

New Discoveries and Lessons Learned

Talking about our experience with colleagues made us realize that what we did is distinctive. We offer our post hoc analysis of success factors rooted in insights from behavioral science in the hope that this inspires and benefits others embarking on a similar mission. The general approach of collaborating on cumulative theory building offers an alternative to the more common competition that can easily cause people merely to challenge each other's efforts, overlooking compatibility and synergetic potential of different research lines. Our experience that adversarial alignment can be fruitful suggests a constructive way forward that may also apply to theoretical debate in other areas of science.

As a side benefit, this kind of collaboration can enhance scientific credibility, earning the respect and trust of nonspecialists in the ability of science not only to establish facts but also, at least to align or even to synthesize them. Although surveys in the United States and Europe suggest that trust of the general public in science is high and stable over time (48, 49), there is also cause for concern. Increased reliance on social media and online reports instead of science outlets and fact-based journalism; corporate and political sponsoring of science; and high-profile incidents that call into question the integrity, autonomy, transparency, and accountability of scientists make it more difficult to communicate the added value of science (50). Examples of adversarial collaboration on research and adversarial alignment of theory can only benefit the achievements and reputation of science.

Data Availability. No data were generated for this manuscript.

1 National Academies of Sciences, Engineering, and Medicine, *Reproducibility and Replicability in Science* (The National Academies Press, Washington, DC, 2019).
2 T. F. Pettigrew, The emergence of contextual social psychology. *Pers. Soc. Psychol. Bull.* **44**, 963–971 (2018).
3 K. Sassenberg, L. Ditrich, Research in social psychology changed between 2011 and 2016: Larger sample sizes, more self-report measures, and more online studies. *Adv. Methods Pract. Psychol. Sci.* **2**, 107–114 (2019).
4 K. Fiedler, The creative cycle and the growth of psychological science. *Perspect. Psychol. Sci.* **13**, 433–438 (2018).
5 M. Muthukrishna, J. Henrich, A problem in theory. *Nat. Hum. Behav.* **3**, 221–229 (2019).
6 W. Mischel, The toothbrush problem. *APS Obs.* **21**, 11 (2008).
7 N. Ellemers, Connecting the dots: Mobilizing theory to reveal the big picture in social psychology (and why we should do this). *Eur. J. Soc. Psychol.* **43**, 1–8 (2013).
8 K. Fiedler, Tools, toys, truisms, and theories: Some thoughts on the creative cycle of theory formation. *Pers. Soc. Psychol. Rev.* **8**, 123–131 (2004).
9 S. T. Fiske, "Building bridges inside and outside social psychology: A case for lumping–neatly" in *Bridging Social Psychology: Benefits of Transdisciplinary Approaches*, P. A. M. Van Lange, Ed. (Lawrence Erlbaum, Mahwah, NJ, 2006), pp. 47–52.
10 S. T. Fiske, *Social Beings: Core Motives in Social Psychology* (Wiley, Hoboken, NJ, ed. 2, 2010).
11 E. T. Higgins, Making a theory useful: Lessons handed down. *Pers. Soc. Psychol. Rev.* **8**, 138–145 (2004).
12 A. W. Kruglanski, "Theories as bridges" in *Bridging Social Psychology: Benefits of Transdisciplinary Approaches*, P. A. M. Van Lange, Ed. (Lawrence Erlbaum, Mahwah, NJ, 2008), pp. 21–34.
13 H. T. Reis, Steps toward the ripening of relationship science. *Pers. Relatsh.* **14**, 1–23 (2007).
14 J. Saatsi, What is theoretical progress of science? *Synthese* **169**, 611–631 (2019).
15 P. A. M. Van Lange, "Bridging social psychology" in *Bridging Social Psychology: Benefits of Transdisciplinary Approaches*, P. A. M. Van Lange, Ed. (Lawrence Erlbaum, Mahwah, NJ, 2006), pp. 3–20.

16 K. D. Martin, J. B. Cullen, Continuities and extensions of ethical climate theory: A meta-analytic review. *J. Bus. Ethics* **69**, 175–194 (2006).
17 A. Simha, J. B. Cullen, Ethical climates and their effects on organizational outcomes: Implications from the past and prophecies for the future. *Acad. Manage. Perspect.* **26**, 20–34 (2012).
18 L. K. Treviño, N. A. den Nieuwenboer, J. J. Kish-Gephart, (Un)ethical behavior in organizations. *Annu. Rev. Psychol.* **65**, 635–660 (2014).
19 K. Tzini, K. Jain, Unethical behavior under relative performance evaluation: Evidence and remedy. *Hum. Resour. Manage.* **57**, 1399–1413 (2018).
20 M. A. Edwards, S. Roy, Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. *Environ. Eng. Sci.* **34**, 51–61 (2017).
21 T. L. Haven, J. K. Tijdink, B. C. Martinson, L. M. Bouter, Perceptions of research integrity climate differ between academic ranks and disciplinary fields: Results from a survey among academic researchers in Amsterdam. *PLoS One* **14**, e0210599 (2019).
22 U. Sandström, P. Van den Besselaar, Funding, evaluation, and the performance of national research systems. *J. Informetrics* **2**, 365–384 (2018).
23 J. B. Ruscher, S. T. Fiske, H. Miki, S. Van Manen, Individuating processes in competition: Interpersonal versus intergroup. *Pers. Soc. Psychol. Bull.* **17**, 595–605 (1991).
24 S. T. Fiske, Stereotype content: Warmth and competence endure. *Curr. Dir. Psychol. Sci.* **27**, 67–73 (2018).
25 M. J. Hornsey, S. Esposo, Resistance to group criticism and recommendations for change: Lessons from the intergroup sensitivity effect. *Soc. Personal. Psychol. Compass* **3**, 275–291 (2009).
26 J. Jetten, M. J. Hornsey, Deviance and dissent in groups. *Annu. Rev. Psychol.* **65**, 461–485 (2014).
27 M. Lamont, *How Professors Think: Inside the Curious World of Academic Judgment* (Harvard University Press, Cambridge, MA, 2009).
28 A. Koch, R. Imhoff, R. Dotsch, C. Unkelbach, H. Alves, The ABC of stereotypes about groups: Agency/socioeconomic success, conservative-progressive beliefs, and communion. *J. Pers. Soc. Psychol.* **110**, 675–709 (2016).
29 B. A. De Jong, K. T. Dirks, N. Gillespie, Trust and team performance: A meta-analysis of main effects, moderators, and covariates. *J. Appl. Psychol.* **101**, 1134–1150 (2016).
30 R. Shazi, N. Gillespie, J. Steen, Trust as a predictor of innovation network ties on project teams. *Int. J. Proj. Manag.* **33**, 81–91 (2015).
31 D. Kahneman, Experiences of collaborative research. *Am. Psychol.* **58**, 723–730 (2003).
32 R. Fisher, W. Ury, B. Patton, *Getting to Yes: Negotiating Agreement without Giving in* (Penguin Books, 2011).
33 W. Steinel, A. E. Abele, C. W. DeDreu, Effects of experience and advice on process and performance in negotiations. *Group Process. Intergroup Relat.* **10**, 533–550 (2007).
34 A. Guinote, How power affects people: Activating, wanting, and goal seeking. *Annu. Rev. Psychol.* **68**, 353–381 (2017).
35 F. Harinck, N. Ellemers, "How values change a conflict" in *Conflicts within and between Groups: Functions, Dynamics, and Interventions*, C. K. W. de Dreu, Ed. (Current Issues in Social Psychology, Psychology Press, London, UK, 2014), pp. 19–36.
36 E. Aronson, N. Blaney, C. Stephan, J. Sikes, M. Snapp, *The Jigsaw Classroom* (Sage Publications, Beverly Hills, CA, 1978).
37 P. M. Gollwitzer, P. Sheeran, Implementation intentions and goal achievement: A meta analysis of effects and processes. *Adv. Exp. Soc. Psychol.* **38**, 69–119 (2006).
38 N. Ellemers, F. Rink, Diversity in work groups. *Curr. Opin. Psychol.* **11**, 49–53 (2016).
39 S. Schulz-Hardt, M. Jochims, D. Frey, Productive conflict in group decision making: Genuine and contrived dissent as strategies to counteract biased information seeking. *Organ. Behav. Hum. Decis. Process.* **88**, 563–586 (2002).
40 R. Brown, M. Hewstone, An integrative theory of intergroup contact. *Adv. Exp. Soc. Psychol.* **37**, 255–343 (2005).
41 T. F. Pettigrew, L. R. Tropp, A meta-analytic test of intergroup contact theory. *J. Pers. Soc. Psychol.* **90**, 751–783 (2006).
42 N. Ellemers, D. De Gilder, S. A. Haslam, Motivating individuals and groups at work: A social identity perspective on leadership and group performance. *Acad. Manage. Rev.* **29**, 459–478 (2004).
43 S. Gaertner, J. Dovidio, *Reducing Intergroup Bias: The Common Ingroup Identity Model* (Psychology Press, New York, NY, 2014).
44 M. Uchronski, A. E. Abele, S. Bruckmüller, The situational malleability of the communal self-concept: How perspective-taking affects self-descriptions. *Self. Ident.* **12**, 236–256 (2012).
45 S. Wening, N. Keith, A. E. Abele, High construal level can help negotiators to reach integrative agreements: The role of information exchange and judgement accuracy. *Br. J. Soc. Psychol.* **55**, 206–226 (2016).
46 J. Z. Rubin, F. E. A. Sander, When should we use agents? Direct vs. representative negotiation. *Negotiation J.* **4**, 395–401 (1988).
47 H. Stanislaw, N. Todorov, Calculation of signal detection theory measures. *Behav. Res. Methods Instrum. Comput.* **31**, 137–149 (1999).
48 C. Funk, B. Kennedy, Public confidence in scientists has remained stable for decades. Pew Research Center (2019). https://www.pewresearch.org/fact-tank/2019/03/22/public-confidence-in-scientists-has-remained-stable-for-decades/. Accessed 26 February 2020.
49 N. Van den Broek-Honingh, J. De Jonge, Trust in science in the Netherlands—survey monitor 2018. The Hague: Rathenau Instituut (2018). https://www.rathenau.nl/en/knowledge-ecosystems/trust-science-netherlands. Accessed 26 February 2020.
50 ALLEA, Trust in science and changing landscapes of communication. Discussion paper, No. 3, January (2019). https://www.allea.org/trust-in-science-in-a-digital-world-new-mechanisms-needed/. Accessed 26 February 2020.
51 A. E. Abele, B. Wojciszke, Agency and communion from the perspective of self versus others. *J. Pers. Soc. Psychol.* **93**, 751–763 (2007).
52 A. E. Abele, B. Wojciszke, Communal and agentic content. A dual perspective model. *Adv. Exp. Soc. Psychol.* **50**, 195–255 (2014).
53 N. Ellemers, S. Pagliaro, M. Barreto, Morality and behavioural regulation in groups: A social identity approach. *Eur. Rev. Soc. Psychol.* **24**, 160–193 (2013).
54 N. Ellemers, *Morality and the Regulation of Social Behavior* (Routledge/Taylor & Francis, Milton Park, UK, 2017).
55 V. Y. Yzerbyt, V. Provost, O. Corneille, Not competent but warm... Really? Compensatory stereotypes in the French-speaking world. *Group Process. Intergroup Relat.* **8**, 291–308 (2005).
56 V. Y. Yzerbyt, "The dimensional compensation model: Reality and strategic constraints on warmth and competence in intergroup perceptions" in *The Agency-Communion Framework*, A. E. Abele, B. Wojciszke, Eds. (Routledge, London, UK, 2018), pp. 126–141.
57 S. T. Fiske, A. J. C. Cuddy, P. Glick, J. Xu, A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *J. Pers. Soc. Psychol.* **82**, 878–902 (2002).
58 A. Koch, R. Imhoff, "Rethinking the nature and relation of fundamental dimensions of meaning" in *Agency and Communion in Social Psychology*, A. Abele, B. Wojciszke, Eds. (Routledge, Abingdon, UK, 2018), pp. 167–179.
59 A. Abele, N. Ellemers, S. Fiske, A. Koch, V. Yzerbyt, Navigating the social world: Shared horizontal and vertical evaluative dimensions. https://doi.org/10.31234/osf.io/b5nq6 (27 February 2020).