Complicated problems, such as violent conflicts, inequality, mass migration, climate change, vaccine hesitancy, and global pandemics, necessitate extraordinary efforts to resolve. As a prominent example, the novel coronavirus, or COVID-19, has affected the lives of almost all human beings, infecting and killing millions (World Health Organization [WHO], n.d.) and severely damaging the world economy (Ayittey et al., 2020). COVID-19 was found to be highly contagious and transmitted from asymptomatic infected individuals (Nishiura et al., 2020; Rothe et al., 2020). Thus, epidemiologists argued that curbing the pandemic would be challenging without the development of an effective vaccine (Chen et al., 2020; Sah et al., 2021), which led to a race to develop one; the first vaccines were approved in December 2020 (e.g., Baden et al., 2021; Callaway, 2020; Polack et al., 2020). This race included dozens of labs around the world that developed and tested potential vaccines in parallel by using their own theories, outcome measures, and benchmarks and focusing on whether their approach was effective and why (Callaway, 2020; Le et al., 2020). At the same time, given the urgency of the situation, many researchers and policymakers pushed for a collaborative approach. Correspondingly, on April 9, 2020, WHO (2020) published a call for a single, large, and international vaccine “tournament” of the vaccines developed at different labs to identify the most effective vaccine compared with a single placebo condition. This call is a useful example of the alternative, result-oriented approach that has gained prominence in the study of medicine to address pressing, complicated, and costly health-related problems (see...
Adaptive Platform Trials Coalition, 2019; Parmar et al., 2014; Wason et al., 2016), although, in the case of the COVID-19 vaccine, it was eventually not endorsed by the research community. We argue that in the social sciences, however, this approach has been underused.

Thus, the aim of the current article is to make the case for and describe the utility of an “intervention tournament” in the psychological sciences to address societal problems. Intervention tournaments (which are sometimes also called “multiarm trials,” “intervention contests,” “comparative evaluations,” or “megastudies”) are studies that compare several different interventions against a single control condition with the same standardized outcome measure(s) and participants drawn from the same population (Bruneau et al., 2018; Efentakis & Politis, 2006; Lai et al., 2014; Milkman et al., 2021; Parmar et al., 2014; for elaborate extensions of this approach used in medical research, see Adaptive Platform Trials Coalition, 2019; Wason et al., 2016). We begin by identifying the two main approaches used to advance solutions and, in particular, interventions to pressing social and medical problems. Then, we focus on intervention tournaments by providing its definition and main types. This is followed by a nuanced discussion that includes some practical considerations and recommendations and the benefits and limitations for conducting psychological intervention tournaments in the lab and field settings to address societal problems.

Two Approaches to Intervention Development

The above example of how researchers, clinicians, practitioners, and policymakers addressed the COVID-19 vaccine development suggests that there are two different approaches in which pressing problems can be addressed, which we refer to as “top-down” and “bottom-up” for sake of simplicity. In contrast to the bottom-up approach that can be done with intervention tournaments, on which we elaborate in the following, the top-down approach—sometimes called the “mechanism-insertion design” (Lai et al., 2014) or “parallel-group randomized control trials” (Adaptive Platform Trials Coalition, 2019)—is the more standard, wildly used, “status quo” approach and has generally been considered the “gold standard” approach (e.g., Adaptive Platform Trials Coalition, 2019; Concato et al., 2000; Kratochwill & Levin, 2014; Slade & Priebe, 2001; Wilson & Juarez, 2015). It entails the development and assessment of an intervention that is grounded in a specific theoretical approach, and its goal is to assess if a specific intervention works and why.

The development of this approach generally progresses in the following manner. First, researchers develop a theory to explain a phenomenon. For example, research that is based on the current coronavirus, as well as on previous SARS and MERS outbreaks, found that one of the main reasons why the coronavirus is highly contagious is because of its spikes (which also gives the virus its name, corona, Latin for “crown”), which can attach to particular proteins in human airway cells (Li, 2016; Tian et al., 2020). Second, following this molecular understanding of the disease, researchers and practitioners then developed an intervention according to their theoretical understanding of coronavirus biology. For example, given the knowledge of the coronavirus spikes, some labs tried to develop vaccines aimed at exposing the body to a spike protein, which would cause the immune system to recognize it as an antigen, or foreign entity (Callaway, 2020; Le et al., 2020; Liu et al., 2020). Third, researchers and practitioners tested this intervention in isolation compared with a control (placebo) condition and in some cases to a second intervention that was previously found to be effective. For example, when developing a vaccine, there is an agreed protocol (with some variants; see e.g., Adaptive Platform Trials Coalition, 2019; Bothwell et al., 2016) that needs to be followed closely to prove a vaccine is effective and safe for use by the general public. This approach includes random assignment between the treatment and placebo groups.

We would argue that although the process of developing most psychological interventions is not identical, it is similar. For example, recent research conducted by Bruneau and colleagues (Kteily et al., 2016; Moore-Berg, Ankori-Karlinsky, et al., 2020; Moore-Berg, Hameiri, & Bruneau, 2020; see also Lees & Cikara, 2020, 2021; Ruggeri et al., 2021) identified that overly pessimistic metaperceptions—or how one thinks out-group members view one’s in-group—are prominent psychological factors that feed intergroup hostility. Specifically, individuals tend to think that adversarial out-group members hold much more negative views of the in-group than they do in reality. For example, Moore-Berg, Ankori-Karlinsky, et al. (2020) found that Democrats and Republicans in the United States think that people from the out-group party dislike and dehumanize them at least twice as much as they actually do, which is strongly associated with antipathy and spiteful policy support that comes at the expense of the country. Given this theoretical understanding, several different interventions that aim to correct overly pessimistic metaperceptions were developed and assessed vis-à-vis a control group (Kteily et al., 2016; Lees & Cikara, 2020; Mernyk et al., 2022). For example, in one study in the...
context of partisan polarization in the United States, Lees and Cikara (2020) developed an intervention in which they showed participants the true value of the partisan out-group perceptions together with the participants’ perceived metaperceptions. Compared with a control condition in which participants were just reminded of their own metaperceptions, the intervention successfully reduced negative metaperceptions and, consequently, negative motivational attributions toward the out-group. This original study was then replicated in nine additional countries, which established the robustness of this psychological intervention (Ruggeri et al., 2021).

Although the benefits of this top-down approach should not be discarded, we argue that it does entail several drawbacks. First, when such studies are conducted, the ability to compare interventions across different studies is challenging. That is, it is unclear whether different studies that compare target interventions with different control groups are comparable, which makes it challenging to determine whether a given intervention is more (or less) effective compared with other interventions. Second, top-down mechanism-in-isolation studies do not necessarily rely on standardized outcome measures and often are examined at different times with different outcome measures. Thus, it is difficult to determine what the most relevant outcomes are and whether these same outcomes would be equally affected by other interventions that aim to ultimately achieve the same goal. These outcomes are based on assumptions and decisions made by the researchers themselves that in some cases could be contested and challenged (Paluck et al., 2019; Pettigrew & Hewstone, 2017). This is exacerbated by what Pettigrew and Hewstone (2017) termed as the “single factor fallacy,” which is the tendency for researchers to rely on their own work, based on their specific theoretical framework and variables they focus on, to develop and assess interventions. As a consequence, important factors, including alternative explanations and theories, or other important variables, may be overlooked. Finally, and as mentioned above, the top-down approach is rather limited in its ability to tackle complicated problems quickly and efficiently because it is based on the available resources each lab has and each lab compares the intervention or interventions with different and unstandardized control groups (e.g., Adaptive Platform Trials Coalition, 2019; Bothwell et al., 2016; Wilson & Juarez, 2015).

Conversely, the bottom-up approach is result-oriented and focuses on finding a solution to a problem by identifying what is effective in the most cost-effective and efficient manner possible with a set of agreed-on outcome measures. This can be done with intervention tournaments. An intervention tournament compares several different interventions against the same control and standardized outcome measures. Interventions can be selected using different criteria (e.g., established theory), as we elaborate on below. The interventions are then assessed with participants drawn from the same population to isolate which intervention or interventions is most effective. The main goal of an intervention tournament is to identify the most successful approach or approaches for mitigating the problem at hand, which means that establishing the mechanism through which one intervention is effective is secondary to the main goal of this approach. Although this approach has gained prominence in medical research, as the WHO (2020) call for a single COVID-19 vaccine tournament exemplifies, to the best of our knowledge and as mentioned above, in the social sciences, this approach is still relatively underused (but see Axelrod & Hamilton, 1981; Bruneau et al., 2018; Lai et al., 2014, 2016; Milkman et al., 2021). Thus, with the aim of facilitating the use of intervention tournaments in the social sciences, in the heart of the current article, we define this approach, review common features and issues when intervention tournaments are used, identify different types of intervention tournaments, and offer recommendations to promote best practices and avoid potential pitfalls in their design, deployment, and reporting.

**Intervention Tournaments**

Broadly speaking, an intervention tournament is an experiment that tests and evaluates the causal effects of different approaches, or interventions, on a set of outcome measures with participants drawn from the same population (Bruneau et al., 2018; Lai et al., 2014, 2016; Milkman et al., 2021; Parmar et al., 2014). As mentioned above, the goal of an intervention tournament is to identify the most successful approach or interventions against the same control and standardized outcome measures. Interventions can be selected using different criteria (e.g., established theory), as we elaborate on below. The interventions are then assessed with participants drawn from the same population to isolate which intervention or interventions is most effective. The main goal of an intervention tournament is to identify the most successful approach or approaches for mitigating the problem at hand, which means that establishing the mechanism through which one intervention is effective is secondary to the main goal of this approach. Although this approach has gained prominence in medical research, as the WHO (2020) call for a single COVID-19 vaccine tournament exemplifies, to the best of our knowledge and as mentioned above, in the social sciences, this approach is still relatively underused (but see Axelrod & Hamilton, 1981; Bruneau et al., 2018; Lai et al., 2014, 2016; Milkman et al., 2021). Thus, with the aim of facilitating the use of intervention tournaments in the social sciences, in the heart of the current article, we define this approach, review common features and issues when intervention tournaments are used, identify different types of intervention tournaments, and offer recommendations to promote best practices and avoid potential pitfalls in their design, deployment, and reporting.
identify the interventions that are the most effective at reducing implicit racial biases. To do this, Lai and colleagues sent out a call for different labs to propose what they believe to be the most effective intervention to reduce implicit racial bias. They then received 17 interventions in total that were based on a diverse set of hypothesized mechanisms (e.g., exposure to positive exemplars, imagined contact, perspective taking, inducing empathy). Lai and colleagues compared all of these interventions with a no-intervention control condition in an iterative process—researchers who contributed interventions to the tournament were able to modify their intervention on the basis of the results of previous rounds of the tournament (i.e., studies), which then led to greater effectiveness in reducing implicit intergroup bias across the studies (for a similar approach, see Axelrod, 1980a, 1980b; Axelrod & Hamilton, 1981).

Across these four intervention-tournament studies, Lai et al. (2014) found eight interventions to be most effective at reducing implicit racial bias. In a follow-up intervention tournament, Lai et al. (2016) found that all eight successful interventions from Lai et al. (2014) were again effective in reducing implicit racial bias when assessed immediately after the interventions but that these effects did not persist when the participants were reassessed several hours to days later. We return to the notion of replicating original results and the longitudinal aspect of intervention tournaments in the next section.

Note that these crowdsourced intervention tournaments are not limited to researchers and research labs, which will likely submit interventions on the basis of their own work and theoretical framework (see Pettigrew & Hewstone, 2017). Crowdsourced intervention tournaments can also be used to solicit interventions developed by practitioners, media experts, filmmakers, and so on. These individuals can develop potentially effective and engaging interventions that rely on their creativity, expertise, experience, intuition, and contextual knowledge (Bar-Tal & Hameiri, 2020; Bruneau et al., 2022; IJzerman et al., 2020). Furthermore, crowdsourcing interventions from those outside of academia can improve the external validity of the research by incorporating interventions already used in the field.

Curated intervention tournaments have gained increased attention in recent years through the work of Bruneau and colleagues (2018; see also Moore-Berg, Hameiri, Falk, & Bruneau, 2022). In this approach, the researchers themselves curate different interventions that they believe have the potential to mitigate a specific problem. Specifically, the curators can take real-world interventions that others, mostly practitioners, have been using, test them in an intervention tournament, and then identify why they were effective, if
Indeed they were. For example, in Bruneau and colleagues’ work, they included various interviews, documentary segments, and news clips available in mainstream media that they—and the practitioners they consulted—thought would reduce levels of Islamophobia. Thus, the interventions themselves are generally created and disseminated before assessment, whether it is in the mainstream media, social media, or elsewhere, and are developed on the basis of the creators’ intuition of what would constitute an effective intervention. However, although these creators may develop compelling and engaging content, they rarely rigorously test the effectiveness of these materials in achieving the desired outcomes (Davidson, 2017).

For example, Bruneau et al. (2018) conducted an intervention tournament to identify videos that most effectively reduce the tendency of non-Muslims to collectively blame all Muslims for the actions of individual Muslim extremists. The underlying theoretical assumption is that collective blame increases Islamophobia among non-Muslims by feeding negative and aggressive attitudes and behaviors toward Muslims. Therefore, the aim of this intervention tournament was to find the most effective intervention to short-circuit this tendency. Thus, Bruneau et al. curated videos that were created by both Muslim and non-Muslim practitioners. The researchers chose these videos because they were both compelling and diverse in terms of styles of delivery (didactic, narrative, satire) and their underlying theories (that were mapped onto them a priori by the researchers). Bruneau et al. found that a 2-min video that showed an interview with a Muslim American woman discussing the tendency of non-Muslim Americans to blame all Muslims for terror attacks but not blame Christians for extremism by individual Christians was most effective at reducing collective blame among non-Muslims and, correspondingly, hostility toward Muslims. Bruneau and colleagues then replicated these results in several follow-up studies using a conceptually similar intervention (Bruneau et al., 2018, 2020).

In a supplemental study, Bruneau et al. (2018) asked an independent set of participants to rate the extent to which they thought the videos would be most effective at reducing Islamophobia. Participants underestimated the potential effect of this particular collective blame video and forecasted that it would be much less effective in reducing collective blame of Muslims compared with the obtained effects. This highlights the need to rigorously test the effectiveness of interventions rather than rely solely on the intuition of researchers, laypeople, and the people who create these interventions and materials. Many of the curated interventions Bruneau and colleagues used in their research (see also Gallardo et al., 2022; Moore-Berg, Hameiri, Falk, & Bruneau, 2022) were developed by practitioners to promote their goal to reduce Islamophobia in the United States but were never experimentally tested. The results of these curated intervention tournaments highlight the importance of conducting this rigorous testing.

Finally, in-house-developed intervention tournaments tend to have a somewhat different goal than crowdsourcing and curated intervention tournaments. In this approach, the interventions created for the intervention tournament are in many cases based on a single theory because they are the product of one group of researchers or a single lab. The different interventions in the tournament can be, for example, a series of interventions with different underlying mechanisms but the same goal, or they can be different iterations of the same intervention with the same underlying goal. As an example from outside of the social sciences, researchers might design different iterations of polymer networks of the same drug to determine the most effective drug-delivery system (Efentakis & Politis, 2006). In the social sciences, this can be, for example, manipulating the delivery techniques and media style (e.g., providing guided self-help via face-to-face meetings vs. via email to address eating disorders; Jenkins et al., 2021) of the same source materials or examining the optimal dose of an intervention (e.g., one vs. three sessions of exposure therapy to prevent the development of posttraumatic stress disorder; Maples-Keller et al., 2020).

For example, Bruneau et al. (2022) created a series of 10 different video interventions aimed at promoting more conciliatory views of FARC ex-combatants among non-FARC Colombians. All of the videos were created by the researchers in collaboration with local filmmakers and were based on the same source material, which included interviews with FARC ex-combatants and their non-FARC Colombian neighbors in a rural demobilization camp. The interviews addressed non-FARC Colombians’ misperceptions of FARC ex-combatants, including their willingness or unwillingness to let go of violence and reintegrate into mainstream Colombian society. Thus, the core of all the videos highlighted evidence of the successful coexistence between these demobilized FARC members and their local neighbors. The main variation between the videos was the interviewee (i.e., ex-combatants, non-FARC Colombians, or both) and the order of their appearance (i.e., ex-combatants before non-FARC Colombians or vice versa). Bruneau et al. found that most videos were effective in reducing antigovernment FARC beliefs and attitudes among non-FARC Colombians and increased support for pro-FARC policies and for the peace process between non-FARC and ex-FARC
Colombians. However, one video, which focused on FARC ex-combatants’ unwillingness and ability to change and reintegrate into the Colombian society and included both FARC ex-combatants and non-FARC responses (in this order), was the most effective both immediately and longitudinally approximately 10 to 12 weeks later. In the intervention tournament and in preregistered follow-up studies, in which these results were replicated, Bruneau et al. also provided some insight into the underlying psychological mechanism. It seemed that this video was effective because it changed participants’ belief about FARC ex-combatants’ ability to change (Goldenberg et al., 2018; Halperin et al., 2011), but not affect toward them, and showed an effect on a behavioral measure that can promote peace and reintegration of FARC ex-combatants.

This highlights the (in-house-developed) intervention tournament as a tool to test different iterations of a similar intervention to zero in on the most effective one. In other words, in most cases, the core is similar, whether it is a drug (Efentakis & Politis, 2006) or the psychological content (Bruneau et al., 2022; Jenkins et al., 2021; Maples-Keller et al., 2020; Milkman et al., 2011; Rosler et al., 2021) that is administered, but the delivery is different in each of these interventions.

Note that in-house-developed intervention tournaments are not limited to testing different iterations of the same underlying intervention and can also include interventions that are based on completely different underlying mechanisms, which are tested against each other (e.g., Hameiri et al., 2018; Van Assche et al., 2020; Yokum et al., 2018) and sometimes against their combination (Kim et al., 2021; Moore-Berg, Hameiri, & Bruneau, 2022; Rosler et al., 2021). For example, Yokum et al. (2018) examined the efficacy of different variations of letters that remind Medicare recipients to get the flu vaccine. These letters were based on different psychological mechanisms and included messages with implementation-intention prompts and enhanced active choice. Yokum et al. found that all letters, regardless of their underlying psychological approach, increased vaccination rates compared with the control condition.

**Key Design Insights From Past Intervention-Tournament Research**

Regardless of which intervention tournament type is chosen, there are various design considerations that researchers should take into account when they develop their intervention tournament. These considerations are aggregated across the existing intervention-tournament literature and provide important design insights into what makes intervention tournaments successful.

**Intervention-tournament viability**

Researchers who have used intervention tournaments have taken several approaches to intervention curation and inclusion, as reviewed above, which include crowdsourcing, curating from available materials, or developing interventions in-house. However, before intervention procurement and tournament deployment, the first question that researchers need to address is whether it is all suitable to conduct an intervention tournament to address the research problem at hand. Indeed, the suitability of an intervention tournament depends on the type of tournament in question. In cases of crowdsourced and curated (and to a lesser extent, in-house-developed) intervention tournaments, the main criterion for conducting one is the wealth of existing theories and practices and a collective urgency to address a particular societal problem. In other words, for an intervention tournament to be viable, it needs to address a problem that many feel is important and contain interventions based on prior research (in case of researchers) or developed materials that were (most likely) not empirically examined before (in case of practitioners). It should come as no surprise, then, that all of the examples we provide throughout this article include pressing global challenges such as prejudice, intergroup conflicts, polarization, Islamophobia, vaccine hesitancy, and so on.

We argue that another criterion for a viable intervention tournament is that the problem that is being addressed is challenging and includes overcoming different psychological barriers (e.g., Bar-Tal & Hameiri, 2020; Hornsey & Fielding, 2017). Challenging problems necessitate diverse contributions (which can be crowdsourced or curated) and out-of-the-box thinking that can address the problem at hand from multiple angles (Uhlmann et al., 2019; Van Bavel et al., 2020). Such problems can also increase the motivation of researchers and practitioners to prove that they can come up with the most efficient solution (e.g., Axelrod & Hamilton, 1981; Bennett & Lanning, 2007; Lai et al., 2014, 2016), which can then be put to the test in a crowdsourced intervention tournament. Finally, addressing challenging problems can also benefit from in-house-developed intervention tournaments, especially when there is an approach that shows promise but needs more fine-tuning to, for example, better circumvent psychological barriers and resistance (e.g., Bruneau et al., 2022).

**Soliciting and incentivizing intervention submissions**

One important issue specifically relevant to crowdsourced intervention tournaments is the process of
soliciting and incentivizing intervention submissions. Although, as mentioned above, conducting an intervention tournament that addresses a pressing and challenging societal problem can increase the motivation of researchers and practitioners to take part in an intervention tournament, there are several reasons why people might still be reluctant to submit interventions to crowdsourced tournaments. These mostly include limited time and insufficient motivation and incentives to take part in big-team science (see Forscher et al., 2020; Uhlmann et al., 2019).

Organizers of crowdsourced intervention tournaments have used different approaches to address these potential problems. Some have provided the deserved recognition of the winner or winners of and other participants in the crowdsourced tournament, such as Anatol Rapaport, who won a tournament that aimed to identify the most effective approach in an iterated Prisoner’s Dilemma game (Axelrod, 1980a, 1980b; Axelrod & Hamilton, 1981; granted, providing due credit and recognition is also important in curated intervention tournaments). In other cases, intervention-tournament organizers award prizes (e.g., monetary, authorship) to intervention developers that withstand the selection process and are admitted into the tournament (see Strengthening Democracy Challenge, 2021) or only to the winners and runners-up (e.g., the Netflix Prize; Bennett & Lanning, 2007). In many cases, tournament organizers offer participants authorship on the main publication that results from the intervention tournament (e.g., Lai et al., 2014, 2016; Parmar et al., 2014). Indeed, some provide recognition, prizes, and authorship on academic publications as incentives (see Strengthening Democracy Challenge, 2021).

Moreover, hypothetically, as an additional incentive, it is also possible to provide participants with the opportunity to write a stand-alone article that pertains to their proposed intervention using the intervention-tournament data (see e.g., Leaver, 2019; Parmar et al., 2014). However, in many cases, this is not feasible because the main intervention-tournament article normally includes all interventions and associated data, which means that for such a stand-alone article to be warranted, researchers will need to develop their own novel research questions and hypotheses (e.g., examining moderations that were not previously investigated). Another associated limitation of the intervention-tournament approach is that this common incentive scheme can in fact *decrease* potential participants’ (in particular, researcher participants) motivation to submit what they might perceive as novel interventions, given that the intervention tournament might decrease their chances of publishing a separate publication in a top-tier journal because it will no longer be novel after being published in the intervention-tournament article.

Finally, although a thorough discussion on potential issues that pertain to authorship are beyond the scope of the current article, we find that in many cases, the intervention-tournament organizers are listed as either first or last authors and all other contributors are listed in an a priori agreed-on order (e.g., in order of effectiveness; see, e.g., Contest Study for Reducing Discrimination in Social Judgement, 2021) or alphabetically in between (e.g., Lai et al., 2014, 2016). In agreement with recent calls for collaborative science, it is advisable that to provide the appropriate intellectual credit that is due to each contributor, a thorough and thoughtful contribution statement should be added using, for example, the CRediT taxonomy (McNutt et al., 2018; for a thorough discussion, see Forscher et al., 2020; Uhlmann et al., 2019).

**Inclusion criteria and amount of interventions**

One question that might arise during intervention curation and deployment is how many interventions to include and what the inclusion criteria should be (for a relevant discussion in medical research, see Lee et al., 2019; Stallard et al., 2009). Deciding what to include in an intervention tournament is not an easy task and is influenced by various factors that are not necessarily under the control of the researchers, such as available resources and number of submissions in crowdsourced tournaments. Which interventions to include could be decided on the basis of the expertise, intuition, and knowledge of the field of the researchers themselves (e.g., Bruneau et al., 2022); by consulting practitioners (e.g., Gallardo et al., 2022); or, more rigorously, through a peer-review process (Strengthening Democracy Challenge, 2021).

The most important inclusion criterion is to include only interventions that have an underlying theoretical basis that differentiates one from the other (whether it is in the content or delivery mode) and warrants their inclusion in the tournament (e.g., Bruneau et al., 2018; Gallardo et al., 2022; Lai et al., 2014, 2016; Moore-Berg, Hameiri, Falk, & Bruneau, 2022). Practically, in case of multiple submissions of a similar intervention, crowdsourced intervention-tournament organizers can, for example, ask submitters to develop the intervention collaboratively or select one submitter according to different criteria, such as having previous publications, publications on the intervention, or initial compelling data that the intervention is successful (see Strengthening Democracy Challenge, 2021).
A second important selection criterion is whether the proposed intervention is expected to affect the tournament’s main outcome variable or variables. This expected efficiency can be determined on the basis of previous literature on the intervention, and there is a special emphasis on studies (in most cases, smaller scale studies) that were conducted preferably in the same context, using the same population, and with the same (or comparable) outcome measures (cf. Lee et al., 2019). For example, if a researcher wants to reduce interpartisan animosity in the United States, a strong intervention contender could be Lees and Cikara’s (2020) metaperception-correction intervention described above. This is because Lees and Cikara found their intervention to be effective at reducing a related outcome measure (support for purposeful obstructionism among partisans) in the same context among a similar population—an effect that was later replicated in many contexts across the globe (Ruggeri et al., 2021).

A third criterion, which is also somewhat at odds with the previous one, is the degree of novelty of the intervention. This novelty is vis-à-vis other competing interventions in the intervention tournament or the current state of the art in research and in practice. As mentioned above, intervention tournaments are an efficient way to examine the effectiveness of interventions (e.g., Freidlin et al., 2008). Thus, in some cases, such as when testing interventions in a context that has not been heavily researched (e.g., using a curated intervention tournament) or trying to refine an already promising intervention (e.g., using an in-house-developed intervention tournament), intervention tournaments allow researchers and practitioners to experiment with novel ideas.

In this case, intervention tournaments can test different ideas that have only an intuitive appeal or anecdotal evidence to support their effectiveness that can be developed specifically for the tournament or curated for it (e.g., Bruneau et al., 2018). It can derive interventions that were established and shown to be effective in one context and population with a set of outcome measures (e.g., self-affirmation intervention to increase group-based guilt, tested in the context of the Israeli–Palestinian conflict and Bosnia and Herzegovina; Čehajić-Clancy et al., 2011) and test them in another context and population with a related set of outcome measures (e.g., improving intergroup attitudes following the Paris and Brussels terror attacks; Van Assche et al., 2020). Finally, novelty can be injected by implementing different principles from other literatures (e.g., on attitude change and persuasion to circumvent resistance to conflict resolution) to test different iterations of an already promising intervention to find the most effective iteration of it (Bruneau et al., 2022; see also Efentakis & Politis, 2006).

We argue that although it makes sense to include as many interventions as possible in the tournament, the number of tested interventions should be based on the resources available to run an intervention tournament that is sufficiently statistically powered to detect differences between each of the different interventions and the control condition (see elaboration below). In practice, intervention tournaments typically vary in the number of interventions included. For instance, whereas Lai et al. (2014) included 18 crowdsourced interventions in their initial tournament (including a sham intervention), Van Assche et al. (2020) included three in-house-developed interventions.

### Potential issues with intervention-inclusion criteria

Intervention tournaments can sometimes include interventions that are diverse in terms of the underlying psychological content, modes of delivery, length, level of engagement, and so on. In some cases, these differences are inevitable because they are an inherent part of the intervention. For example, when it comes to reducing implicit prejudice (Lai et al., 2014, 2016), some researchers might argue that the best way to address this issue is by teaching individuals a new skill that can help them respond empathically, which addresses an important aspect of prejudice. Other researchers might argue that to effectively combat prejudice, one has to provide inconsistent information to change people’s attitudes regarding the prejudiced group (see Bar-Tal & Hameiri, 2020; Hameiri et al., 2014). This can be done, for example, by facilitating some form of intergroup contact (e.g., in person, vicarious, or imagined; e.g., Crisp & Turner, 2012; Dovidio et al., 2017; Pettigrew & Tropp, 2006). These types of interventions will undoubtedly be operationalized differently. In one, participants might be asked to participate in a 15-min-long session in which they learn and exercise a new tool that can help them express more empathy (see the work on cognitive reappraisal as an acquired tool to promote better intergroup relations; e.g., Halperin et al., 2014; Hurtado-Parrado et al., 2019). In the other, they might be asked to passively watch—or actively watch when it comes to virtual reality—a short 2- to 5-min video that includes members of the prejudiced out-group (e.g., Bruneau et al., 2018; Hasson et al., 2019).

Granted, these differences are sometimes unavoidable and in fact might provide critical information about the effectiveness of the different interventions as real-world interventions (see e.g., Bar-Tal & Hameiri, 2020). For example, a lengthy intervention might be less successful, which would indicate that although the psychological content has the potential to reduce prejudice,
people lose the motivation to go through a longer intervention, which ultimately renders the intervention to be less effective (e.g., Tamir, 2009; Tamir et al., 2020). In other cases, in which the interventions are curated or created in-house, it is likely that the characteristics (e.g., mode of delivery, length) of the interventions will be easier to control. In these cases, it is advisable to include interventions with similar characteristics to reduce a potential confound. However, these characteristics can also be controlled in a crowdsourced intervention tournament (see e.g., Lai et al., 2014; Strengthening Democracy Challenge, 2021). For example, guidelines can request that all submitted interventions be ethical (e.g., that they do not include any deception), be completed in less than a specific amount of time (e.g., 5 min), be online, and be costless (e.g., that they do not provide any additional monetary incentives for participation).

Although this approach has its merits, because it allows for an intervention tournament that creates an even playing field for the participating interventions (and reduces the risk of potential confounds, as mentioned above), it also points to a limitation in the intervention-tournament approach that should be noted. Although some variation across interventions might be acceptable, most intervention tournaments compare short (or light touch) interventions, which may have the potential to be scaled up relatively easily but might also be less effective, especially across time, than longer and more intense interventions (Paluck et al., 2021), such as months-long in-person or virtual-contact interventions (e.g., Bruneau, Hameiri, et al., 2021; Mousa, 2020). Contact interventions can still be included in an intervention tournament; however, for them to be comparable with other interventions, they are normally parasocial or vicarious (as opposed to in person) contact interventions (e.g., Gallardo et al., 2022). Although longer, more intense interventions can be tested in intervention tournaments, it is much rarer and usually done as part of an in-house-developed intervention tournament and in collaboration with organizations that provide the infrastructure for such a complicated endeavor either in educational, organizational, or clinical settings, in most cases (Jenkins et al., 2021; Leventhal et al., 2015; Maples-Keller et al., 2020). As an example, Leventhal et al. (2015) examined different months-long curricula—including two separate curricula and their combination compared with an active control—to promote resilience and well-being among girls in 76 schools in Bihar, India.

**Deploying intervention tournaments**

Following intervention procurement, the researchers are then tasked with considering how to deploy the intervention tournament. In most cases, a lab setting might be more feasible in terms of available resources. It can also be more useful because it allows researchers to have more control over the experimental design (Wilson et al., 2010), but potentially at the expense of external validity (e.g., Eastwick et al., 2013; Mitchell, 2012). This includes assuring that there will be no spill-over between the conditions, which is more likely to occur when the study includes numerous conditions and all participants are sampled from the same population. Intervention tournaments can also be conducted in field settings. For instance, researchers could partner with organizations that deploy interventions in the community and work with them to develop an intervention tournament with a randomized control design (see e.g., Milkman et al., 2021; Yokum et al., 2018). On top of the clear benefits to external validity, this type of deployment can foster collaborations across both laboratories and practitioners, which creates easier access to testing the intervention in hard-to-reach places (see Acar et al., 2020; Bar-Tal & Hameiri, 2020; Moore-Berg, Bernstein, et al., 2022). However, field intervention tournaments should be conducted with extra care to avoid doing more harm than good, especially when the interventions that are tested do not have a clear track record that establishes their effectiveness (e.g., in curated intervention tournaments).

**Comparing intervention efficiency**

Following the curation and testing of the interventions, researchers are then tasked with deciding how to compare the interventions. There are three important decisions to make: (a) what the outcome measure or measures are, (b) what statistics to compare, and (c) what type of control condition or conditions to include in the intervention tournament. As mentioned above, one of the first decisions the researchers make when considering whether to deploy an intervention tournament is what problem (ranging from concrete to abstract) to address. This decision then directly translates to the specific outcome measures that are included in the tournament. Concrete problems are usually operationalized as one specific outcome measure. For example, Milkman et al. (2021) attempted to increase flu vaccination by using nudging text messages. Therefore, their only outcome measure was the extent to which their participants got vaccinated. Likewise, Lai et al. (2014) attempted to reduce implicit prejudice and therefore focused on participants’ Implicit Association Test (IAT) scores as an outcome measure (although they did include one additional measure of self-reported racial attitudes, the tournament’s winners were decided using only the IAT scores).
Abstract problems, on the other hand, usually mean that different outcome variables are measured that reflect different operationalizations of the abstract problem. These, in many cases, include behavioral and policy-relevant measures (in addition to attitudinal or affective measures) that are more closely related to the problem. For example, Bruneau et al. (2022) aimed to promote peace and reintegration of FARC ex-combatants in Colombia. Therefore, the researchers measured a variety of outcome measures but focused on a few that included, most notably, participants’ support for the peace process in Colombia and their support for policies that aim to help ex-combatants reintegrate into Colombian society. On top of these outcomes, Bruneau et al. also measured relevant attitudinal and affective measures such as the perception that FARC ex-combatants are unwilling and unable to let go of violence, dehumanization, intergroup empathy, and prejudice.

Finally, in between those two ends of the spectrum, there are some instances in which a problem can be operationalized in several different concrete ways. For example, the Strengthening Democracy Challenge (2021) organizers are interested in strengthening U.S. democracy given rising levels of polarization and interpartisan prejudice (e.g., Iyengar et al., 2019). This rather abstract goal was then operationalized as three concrete outcomes (i.e., antidemocratic attitudes, support for partisan violence, and partisan animosity, which includes a behavioral measure). In other words, the tournament organizers were equally interested in promoting better interpartisan relations and combating the process of democratic norm erosion in the United States.

Next, in the vast majority of intervention tournaments, efficiency of interventions is determined by assessing whether the difference between each intervention and the control condition is statistically significant (below we elaborate on whether these analyses should be adjusted because of multiple comparisons) and by reporting the size of these effects. In the minority of cases, following the initial intervention tournament, and as we elaborate below, the efficiency of interventions is then assessed in a follow-up study that examines whether the effects persisted for a period of time and in replication studies that examine whether the effect is reliable (see e.g., Bruneau et al., 2022; Gallardo et al., 2022; Lai et al., 2014, 2016; Moore-Berg, Hameiri, Falk, & Bruneau, 2022).

It is common for researchers to compare the interventions to an empty control (i.e., no intervention) condition rather than to an active control (placebo) condition. This is because it is extremely difficult in the psychological sciences to come up with an active control condition that will not bias the results in any way (e.g., increase positive affect, cognitive flexibility), especially compared with several other, theoretically diverse, competing interventions. However, an empty control condition has two major limitations that an active control condition can address. Specifically, when an empty control is used, participants can realize that they are in the control condition, which can then affect how they respond to the outcome measures. Indeed, this is less of a concern when the outcomes are behavioral (e.g., in the case of getting flu vaccinations; Milkman et al., 2021) or implicit (e.g., in the case of the IAT; Lai et al., 2014, 2016). However, when the outcome measures are mostly self-reported, then the fact that participants are not blind to their condition can yield potential bias in the results because of demand characteristics (for a related discussion in medical research, see Freidlin et al., 2008). Second, using an active control can reduce potential selection bias (albeit, it cannot be completely eliminated; see Uschner et al., 2018). On the other hand, unlike an active control, an empty control prevents a potentially biased control condition and, as noted previously, requires no additional resources to deploy (i.e., does not require additional resources to develop the active control condition) and provides a potentially true baseline of attitudes and opinions.

Addressing Type I and II errors

Regardless of which type of control condition the interventions are compared with, multiple comparisons are being made, which requires the researchers to carefully consider how to address potential Type I (i.e., false positive) and Type II (i.e., false negative) errors. There are diverging views about whether researchers should include some (and what type of) statistical correction because multiple comparisons are being made. Some have argued that no correction is needed, especially in cases of intervention tournaments that examine distinct interventions and is exploratory, because the comparisons that are being made (between each intervention and control) are independent (Parker & Weir, 2020; Rubin, 2021). Others have argued that some type of correction is needed, especially in cases of intervention tournaments that examine iterations of the same intervention (which is mostly relevant to the in-house-developed intervention tournaments) and are confirmatory (Freidlin et al., 2008; Wason et al., 2014; Wason & Robertson, 2021). Note that intervention tournaments in the psychological sciences are more often exploratory than confirmatory.

Given these diverging views, we argue that to rule out potential Type I and Type II statistical errors, it is important for researchers to include multiple, preregistered,
and statistically powered replication studies with new participants (e.g., Bruneau et al., 2022; Calanchini et al., 2021; Lai et al., 2014, 2016; Moore-Berg, Hameiri, Falk, & Bruneau, 2022). For instance, in Moore-Berg, Hameiri, Falk, and Bruneau (2022), the researchers initially conducted a 12-condition intervention tournament to examine the effectiveness of video interventions at reducing Islamophobia. From this initial intervention tournament, the authors identified three interventions that were most successful at decreasing support for punitive policies toward Muslims. They then conducted five follow-up preregistered and large-scale replication studies to rule out potential Type I and II errors. Likewise, Lai et al. (2014) conducted several replication studies following the initial intervention tournament and eventually tested each intervention 3.70 times on average (see also Axelrod, 1980a, 1980b). By replicating the results in preregistered and sufficiently powered studies, researchers can considerably reduce the concern that the effects they find in the intervention tournament are a mere fluke, which increases the results’ robustness. If replication studies are not feasible, researchers are advised to use various statistical techniques to account for multiple comparisons, such as reporting $q$ values on top of the customary statistical indices (see Milkman et al., 2021; Wason & Robertson, 2021).

**Underlying mechanisms of interventions**

In addition to ruling out Type I and II errors, replication studies add further benefit of teasing out the underlying mechanism or mechanisms of the successful interventions identified in the tournament, especially when considering the curated and in-house-developed intervention tournaments. In most cases, researchers and practitioners will have some sense of the potential psychological mechanisms at play before the intervention tournament, as in Bruneau et al. (2018) and Moore-Berg, Hameiri, Falk, and Bruneau (2022; see also Calanchini et al., 2021). In other cases, researchers can tease apart a successful intervention after it is found to be successful to pinpoint the psychological mechanism from potential candidates. Both approaches have unique benefits associated with them and are not mutually exclusive. For instance, focusing on specific mechanisms before the intervention tournament can help the researchers during intervention curation. That is, the researchers might be interested in identifying interventions that increase empathy toward out-group members. Therefore, they might select only videos that appear to induce empathy on the basis of their intuitions. Then, the researchers might conduct a confirmatory study to ensure that empathy was the mechanism involved and/or tease apart additional psychological mechanisms that might be at play. However, there might be other cases in which researchers want to take a more exploratory approach to the intervention tournament and include, for example, interventions generally thought to improve implicit attitudes toward out-group members (see Lai et al., 2014). In this case, the researchers might include an assortment of promising interventions that appear to improve intergroup relations and focus only on identifying the mechanism after determining which intervention or interventions are most effective. For this more exploratory approach, the researchers might consider selecting several theoretical mechanisms that could drive the intervention’s effects a priori and then rule out which theoretical mechanism is most effective in subsequent follow-up studies.

**Longitudinal intervention tournaments**

As a final consideration for intervention-tournament design, some tournaments include a longitudinal component (e.g., Bruneau et al., 2022; Lai et al., 2016; Moore-Berg, Hameiri, Falk, & Bruneau, 2022). Including a longitudinal component as part of intervention-tournament design has several benefits. First, it can serve as another criterion for determining which intervention or interventions are most successful. For instance, Moore-Berg, Hameiri, Falk, and Bruneau (2022) administered the same questionnaire (without the interventions) to the same participants in the intervention tournament 1 month following the tournament. They considered an intervention to be successful only when it maintained its significant effects on the outcome during the 1-month follow-up study (see also Bruneau et al., 2022). Second, a longitudinal component can be another way to rule out Type I error and demand characteristics. By demonstrating that an effect lasts over time, researchers can have increased confidence that the effect was driven by the intervention itself rather than a statistical or methodological errors. Third, it can increase researchers’ confidence in the effectiveness of the intervention. Like all longitudinal studies, demonstrating that an effect lasts beyond the initial testing increases the robustness of the research. Unfortunately, only a small minority of all studies that assess the effectiveness of interventions include a longitudinal element (Paluck et al., 2021).

**Conclusion**

The goal of the current article was to increase the use of intervention tournaments in the psychological sciences by providing the pros and cons of this approach and practical considerations for psychological scientists who are interested in using it. We argue that intervention
tournaments hold the potential to greatly improve scientific research—they allow for the efficiency of testing of multiple interventions at once, encourage the collaboration across research labs with diverse expertise and/or between academics and practitioners, and improve the external validity of research. Intervention tournaments also hold the potential to increase the rigor of applied research that assesses interventions that are designed and implemented by practitioners. However, we note that intervention tournaments are not a panacea. Although it is an efficient approach, because it tests several interventions at once by means of a single control condition, a single study does require considerable resources to ensure sufficient statistical power, especially if researchers want to use a nationally representative sample or collaborate with organizations that have the capabilities and infrastructure to run such complicated studies in the field. Furthermore, and as mentioned above, the focus on identifying the more effective approach means that the intervention tournament does not inherently provide an answer to why the successful intervention or interventions was indeed the most effective.

As we have elaborated on in this article, there are several elements that we suggest researchers should incorporate in their research to use this approach effectively while also minimizing its limitations. These include, most notably, conducting replication studies that can address issues of statistical errors and shed light on potential psychological mechanisms and some limitations in the intervention-selection process. We argue that when this approach is used diligently, intervention tournaments complement existing approaches to intervention science by providing the opportunity for rigorous investigation of interventions against other interventions. This sort of rigorous testing is necessary to push the field of intervention science forward and establish theoretical bases for which interventions are most successful.

**Transparency**

*Action Editor:* Adam Cohen  
*Editor:* Laura A. King  
*Author Contributions:*  
B. Hameiri and S. L. Moore-Berg contributed equally to this work and are listed in alphabetical order.  
*Declaration of Conflicting Interests:*  
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.  
*Funding:*  
This work was supported by Israel Science Foundation (ISF) Grant 1590/20 (awarded to B. Hameiri).

**ORCID iDs**

Boaz Hameiri https://orcid.org/0000-0002-0241-9839  
Samantha L. Moore-Berg https://orcid.org/0000-0003-2972-2288

**Acknowledgments**

We are grateful to Emile Bruneau for inspiration; many of the ideas in this article reflect conversations and collaborations with Emile in his effort to put science to work for peace, and we were deeply saddened by his loss to brain cancer on September 30, 2020. We thank Daniel Bar Tal, Emily Falk, and Michael Pasek for their helpful comments on earlier versions of the article.

**Note**

1. The FARC is a leftist insurgent movement that took up arms in 1964 to protect indigenous and poor communities from exploitation by governmental and business interests in Colombia.

**References**

Acar, Y. G., Moss, S. M., & Ulug, Ö. M. (Eds.). (2020). Researching peace, conflict, and power in the field: Methodological challenges and opportunities. Springer.

Adaptive Platform Trials Coalition. (2019). Adaptive platform trials: Definition, design, conduct and reporting considerations. *Nature Reviews Drug Discovery, 18*(10), 797–808. https://doi.org/10.1038/s41573-019-0034-3

Axelrod, R. (1980a). Effective choice in the prisoner’s dilemma. *Journal of Conflict Resolution, 24*(1), 3–25.

Axelrod, R. (1980b). More effective choice in the prisoner’s dilemma. *Journal of Conflict Resolution, 24*(3), 379–403.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*(4489), 1390–1396.

Ayittey, F. K., Ayittey, M. K., Chiwero, N. B., Kamasah, J. S., & Dzuvor, C. (2020). Economic impacts of Wuhan 2019-nCoV on China and the world. *Journal of Medical Virology, 92*(5), 473–475.

Baden, L. R., El Sahly, H. M., Essink, B., Kotloff, K., Frey, S., Novak, R., Diemert, D., Spector, S. A., Rouphael, N., Creech, C. B., McGettigan, J., Khetan, S., Segall, N., Solis, J., Brosz, A., Fierro, C., Schwartz, H., Neuzil, K., Corey, L., . . . COVE Study Group. (2021). Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *The New England Journal of Medicine, 384*(5), 403–416. https://doi.org/10.1056/nejmoa2035389

Bar-Tal, D., & Hameiri, B. (2020). Interventions to change well-anchored attitudes in the context of intergroup conflict. *Social and Personality Psychology Compass, 14*(7), Article e12534. https://doi.org/10.1111/spc3.12534

Bennett, J., & Lanning, S. (2007). The Netflix prize. *Proceedings of the KDD Cup and Workshop*. https://www.cs.uic.edu/~liub/KDD-cup-2007/proceedings.html

Bothwell, L. E., Greene, J. A., Podolsky, S. H., & Jones, D. S. (2016). Assessing the gold standard—Lessons from the history of RCTs. *The New England Journal of Medicine, 374*, 2175–2181.

Bruneau, E. (2015). Putting neuroscience to work for peace. In E. Halperin & K. Sharvit (Eds.), *The social psychology of intractable conflict: Celebrating the legacy of Daniel Bar-Tal* (Vol. 1, pp. 143–155). Springer.
Bruneau, E., Casas, A., Hameiri, B., & Kteily, N. (2022). Exposure to a media intervention helps promote peace and reintegration in Colombia. *Nature Human Behaviour*. Advance online publication. https://doi.org/10.1038/s41562-022-01350-w

Bruneau, E., Hameiri, B., Moore-Berg, S. L., & Kteily, N. (2021). Intergroup contact reduces dehumanization and meta-dehumanization: Cross-sectional, longitudinal, and quasi-experimental evidence from 16 samples in five countries. *Personality and Social Psychology Bulletin, 47*, 906–920. https://doi.org/10.1177/014616720949004

Bruneau, E., Kteily, N., & Falk, E. (2018). Interventions highlighting hypocrisy reduce collective blame of Muslims for individual acts of violence and assuage anti-Muslim hostility. *Personality and Social Psychology Bulletin, 44*, 430–448.

Bruneau, E. G., Kteily, N. S., & Urbiola, A. (2020). A collective blame hypocrisy intervention enduringly reduces hostility towards Muslims. *Nature Human Behaviour, 4*, 45–54.

Calanchini, J., Lai, C. K., & Klauer, K. C. (2021). Reducing implicit racial preferences: III. A process-level examination of changes in implicit preferences. *Journal of Personality and Social Psychology, 121*(4), 796–818. https://doi.org/10.1037/pspi0000339

Callaway, E. (2020). The race for the coronavirus vaccines. *Nature, 580*, 576–577.

Čechajíc-Clancy, S., Effron, D. A., Halperin, E., Liberman, V., & Ross, L. D. (2011). Affirmation, acknowledgment of in-group responsibility, group-based guilt, and support for reparative measures. *Journal of Personality and Social Psychology, 101*(2), 256–270.

Chen, W.-H., Strych, U., Hotez, P. J., & Bottazzi, M. E. (2020). The SARS-CoV-2 vaccine pipeline: An overview. *Current Tropical Medicine Reports, 7*, 61–64.

Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *The New England Journal of Medicine, 372*(9), 793–795.

Concato, J., Shah, N., & Horwitz, R. I. (2000). Randomized, controlled trials, observational studies, and the hierarchy of research designs. *The New England Journal of Medicine, 342*, 1887–1892.

Contest Study for Reducing Discrimination in Social Judgement. (2021). *Contest study for reducing discrimination in social judgement: The collaborator's guide*. https://drive.google.com/file/d/1V0XZN2H-UX3S-v2doMBnKrfxWgaShwEx/view

Crisp, R. J., & Turner, R. N. (2012). The imagined contact hypothesis. In J. M. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (Vol. 46, pp. 125–182). Academic Press.

Cuijpers, P., Ebert, D. D., Acarturk, C., Andersson, G., & Cristea, I. A. (2016). Personalized psychotherapy for adult depression: A meta-analytic review. *Behavior Therapy, 47*, 966–980.

Davidson, B. (2017). Storytelling and evidence-based policy: Lessons from the grey literature. *Palgrave Communications, 3*, 1–10.

Dovidio, J. F., Love, A., Schellhaas, F. M. H., & Hewstone, M. (2017). Reducing intergroup bias through intergroup contact: Twenty years of progress and future directions. *Group Processes & Intergroup Relations, 20*, 606–620.

Eastwick, P. W., Hunt, L. L., & Neff, L. A. (2013). External validity, why art thou externally valid? Recent studies of attraction provide three theoretical answers. *Social and Personality Psychology Compass, 7*(5), 275–288.

Efentakis, M., & Politis, S. (2006). Comparative evaluation of various structures in polymer controlled drug delivery systems and the effect of their morphology and characteristics on drug release. *European Polymer Journal, 42*, 1185–1195.

Forscher, P. S., Wagemakers, E., Coles, N. A., Silan, M. A., Dutra, N. B., Basnight-Brown, D., & IJzerman, H. (2020). *The benefits, barriers, and risks of big team science*. OSF. https://doi.org/10.31234/osf.io/2mdxh

Freidlin, B., Korn, E. L., Gray, R., & Martin, A. (2008). Multi-arm clinical trials of new agents: Some design considerations. *Clinical Cancer Research, 14*, 4368–4371. https://doi.org/10.1158/1078-0432.CCR-08-0325

Gallardo, R. A., Hameiri, B., & Moore-Berg, S. L. (2022). American-Muslims’ use of humor to defuse hate speech as an effective anti-Islamophobia intervention [Manuscript submitted for publication]. Annenberg School for Communication, University of Pennsylvania.

Goldenberg, A., Cohen-Chen, S., Goyer, J. P., Dweck, C. S., Gross, J. J., & Halperin, E. (2018). Testing the impact and durability of group malleability intervention in the context of the Israeli-Palestinian conflict. *Proceedings of the National Academy of Sciences, USA, 115*, 696–701.

Halperin, E., Pliskin, R., Saguy, T., Liberman, V., & Gross, J. J. (2014). Emotion regulation and the cultivation of political tolerance: Searching for a new track for intervention. *Journal of Conflict Resolution, 58*, 1110–1138.

Halperin, E., Russell, A. G., Trzesniewski, K. H., Gross, J. J., & Dweck, C. S. (2011). Promoting the Middle East peace process by changing beliefs about group malleability. *Science, 333*, 1767–1769.

Halperin, E., & Schori-Eyal, N. (2020). Towards a new framework of personalized psychological interventions to improve intergroup relations and promote peace. *Social and Personality Psychology Compass, 14*, 255–270.

Hameiri, B., Bar-Tal, D., & Halperin, E. (2014). Challenges for peacemakers: How to overcome socio-psychological barriers. *Policy Insights from the Behavioral and Brain Sciences, 1*, 164–171.

Hameiri, B., Nabet, E., Bar-Tal, D., & Halperin, E. (2018). Paradoxical thinking as a conflict-resolution intervention: Comparison to alternative interventions and examination of psychological mechanisms. *Personality and Social Psychology Bulletin, 44*, 122–139.

Hasson, Y., Schori-Eyal, N., Landau, D., Hasler, B. S., Levy, J., Friedman, D., & Halperin, E. (2019). The enemy’s gaze: Immersive virtual environments enhance peace promoting attitudes and emotions in violent intergroup conflicts. *PLOS ONE, 14*, Article e0222342. https://doi.org/10.1371/journal.pone.0222342

Hirsh, J. B., Kang, S. K., & Bodenhausen, G. V. (2012). Personalized persuasion: Tailoring persuasive appeals to recipients’ personality traits. *Psychological Science, 23*(6), 578–581. https://doi.org/10.1177/0956797611436349
Hornsey, M. J., & Fielding, K. S. (2017). Attitude roots and Jiu Jitsu persuasion: Understanding and overcoming the motivated rejection of science. *American Psychologist*, 72, 459–473.

Hurtado-Parrado, C., Sierra-Puentes, M., El Hazzouri, M., Morales, A., Gutiérrez-Villamarín, D., Velásquez, L., Correa-Chica, A., Rincón, J. C., Henao, K., Castañeda, J. G., & López-López, W. (2019). Emotion regulation and attitudes toward conflict in Colombia: Effects of reappraisal training on negative emotions and support for conciliatory and aggressive statements. *Frontiers in Psychology*, 10, Article 908. https://doi.org/10.3389/fpsyg.2019.00908

Ijzerman, H., Lewis, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., Vazire, S., Forscher, P. S., Morey, R. D., Ivory, J. D., & Anvari, F. (2020). Use caution when applying behavioural science to policy. *Nature Human Behaviour*, 4, 1092–1094. https://doi.org/10.1038/s41562-020-00990-w

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22(1), 129–146.

Jenkins, P. E., Luck, A., Violato, M., Robinson, C., & Fairborn, Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Fielding, K. S. (2017). Attitude roots and Jiu Jitsu persuasion: Understanding and overcoming the motivated rejection of science. *American Psychologist*, 72, 459–473.

Kim, S., Richardson, A., Werner, P., & Anstey, K. J. (2021). Dementia stigma reduction (DESeRvE) through education and virtual contact in the general public: A multi-arm factorial randomised controlled trial. *Dementia*, 20, 2152–2169.

Kratochwill, T. R., & Levin, J. R. (Eds.). (2014). *School psychology series. Single-case intervention research: Methodological and statistical advances*. American Psychological Association. https://doi.org/10.1037/14376-000

Kteily, N., Hodson, G., & Bruneau, E. (2016). They see us as less than human: Metadehumanization predicts intergroup conflict via reciprocal dehumanization. *Journal of Personality and Social Psychology*, 110, 343–370.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. E. L., Joy-Gaba, J. A., Ho, A. K., Teachman, B. A., Wojcik, S. P., Koleva, S. P., Frazier, R. S., Heiphetz, L., Chen, E. E., Turner, R. N., Haidt, J., Kesebir, S., Hawkins, C. B., Schaefer, H. S., Rubichi, S., . . . Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, 143, 1765–1785. https://doi.org/10.1037/a0036260

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., Calanchini, J., Xiao, Y. J., Pedram, C., Marshburn, C. K., Simon, S., Blanchard, J. C., Joy-Gaba, J. A., Conway, J., Redford, L., Klein, R. A., Roussos, G., Schellhaas, F. M., Burns, M., . . . Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, 145, 1001–1016. https://doi.org/10.1037/xge0000179

Le, T. T., Andreadakis, Z., Kumar, A., Román, R. G., Tollefsen, S., Saville, M., & Mayhew, S. (2020). The COVID-19 vaccine development landscape. *Nature Review Drug Discovery*, 19, 305–306.

Leaver, J. (2019). *Metaketa initiative field guide*. Evidence in Governance and Politics. https://egap.org/our-work-0/the-metaketa-initiative/

Lee, K. M., Wason, J., & Stallard, N. (2019). To add or not to add a new treatment arm to a multiarm study: A decision-theoretic framework. *Statistics in Medicine*, 38(18), 3505–3521.

Lees, J., & Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behaviour*, 4, 279–286.

Lees, J., & Cikara, M. (2021). Understanding and combating misperceived polarization. *Philosophical Transactions of the Royal Society B*, 376, Article 20200143. https://doi.org/10.1098/rstb.2020.0143

Leventhal, K. S., DeMaria, L. M., Gillham, J., Andrew, G., Peabody, J. W., & Leventhal, S. (2015). Fostering emotional social, physical and educational wellbeing in rural India: The methods of a multi-arm randomized controlled trial of Girls First. *Trials*, 16, Article 481. https://doi.org/10.1186/s13063-015-1008-3

Li, F. (2016). Structure, function, and evolution of coronavirus spike protein. *Annual Review of Virology*, 3, 237–261.

Liu, C., Zhou, Q., Li, Y., Garner, L. V., Watkins, S. P., Carter, L. J., Smoot, J., Gregg, A. C., Daniels, A. D., Jervey, S., & Albau, D. (2020). Research and development on therapeutic agents and vaccines for COVID-19 and related human coronavirus diseases. *ACS Central Science*, 6, 315–331. https://doi.org/10.1021/acscentsci.0c00272

Maples-Keller, J. L., Post, L. M., Price, M., Goodnight, J. M., Burton, M. S., Yasinski, C. W., Michopoulos, V., Stevens, J. S., Hinrichs, R., Rothbaum, A. O., Hudak, L., Houry, D., Jovanovic, T., Ressler, K., & Rothbaum, B. O. (2020). Investigation of optimal dose of early intervention to prevent posttraumatic stress disorder: A multiarm randomized trial of one and three sessions of modified prolonged exposure. *Depression & Anxiety*, 37, 429–437. https://doi.org/10.1002/da.23015

McNutt, M. K., Bradford, M., Drazen, J. M., Hanson, B., Howard, B., Jamieson, K. H., Kiermer, V., Marcus, E., Pope, B. K., Schekman, R., Swaminathan, S., Stang, P. J., & Verma, I. M. (2018). Transparency in authors’ contributions and responsibilities to promote integrity in scientific publication. *Proceedings of the National Academy of Sciences, USA*, 115(11), 2557–2560. https://doi.org/10.1073/pnas.1715374115

Mernyk, J. S., Nosek, B. A., & Verma, I. M. (2018). Transparency in authors' contributions and responsibilities to promote integrity in scientific publication. *Proceedings of the National Academy of Sciences, USA*, 115(11), 2557–2560. https://doi.org/10.1073/pnas.1715374115

Mernyk, J. S., Pink, S. L., Druckman, J. N., & Willer, R. (2022). Correcting inaccurate metaperceptions reduces Americans’ support for partisan violence. *Proceedings of the National Academy of Sciences, USA*, 119(16), Article e2116851119. https://doi.org/10.1073/pnas.2116851119

Milkan, K. L., Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2011). Using implementation intentions prompts to enhance influenza vaccination rates. *Proceedings of the National Academy of Sciences, USA*, 108, 10415–10420.
Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice reduction: Progress and challenges. Annual Review of Psychology, 72, 533–560.

Parker, R. A., & Weir, C. J. (2020). Non-adjustment for multiple testing in multi-arm trials of distinct treatments: Rationale and justification. Clinical Trials, 17(5), 562–566.

Parmar, M. K. B., Carpenter, J., & Sydes, M. R. (2014). More multiarm randomized trials of superiority are needed. The Lancet, 384, 283–284.

Pettigrew, T. F., & Hewstone, M. (2017). The single factor fallacy: Implications of missing critical variables from an analysis of intergroup contact theory. Social Issues and Policy Review, 11, 8–37.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. Journal of Personality and Social Psychology, 90, 751–783.

Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J. L., Perez Marc, G., Moreira, E. D., Zerbini, C., Bailey, R., Swanson, K. A., Roychoudhury, S., Koury, K., Li, P., Kalina, W. V., Cooper, D., French, R. W., Jr., Hammitt, L. L., . . . C4591001 Clinical Trial Group. (2020). Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. The New England Journal of Medicine, 383, 2603–2615. https://doi.org/10.1056/NEJMoa2034577

Rosler, N., Sharvit, K., Hameiri, B., Weiner-Blotner, O., Idan, O., & Bar-Tal, D. (2021). The informative process model as a new intervention for attitude change in intractable conflicts [Manuscript submitted for publication]. Program in Conflict Resolution and Mediation, Tel Aviv University.

Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froschkl, G., Wallrauch, C., Zimmer, T., Thiell, V., Janke, C., Guggemos, W., Selmaier, M., Drosten, C., Vollmar, P., Zwirglmaier, K., Zange, S., Wölfel, R., & Hoelscher, M. (2020). Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. The New England Journal of Medicine, 382, 970–971. https://doi.org/10.1056/NEJMc2001468

Rubin, M. (2021). When to adjust alpha during multiple testing: A consideration of disjunction, conjunction, and individual testing. Synthese, 199, 10969–11000. https://doi.org/10.1007/s11229-021-03276-4

Ruggeri, K., Večkalov, B., Bojanić, L., Andersen, T. L., Ashcroft-Jones, S., Ayacaxli, N., Barca-Arroyo, P., Berge, M. L., Bjørndal, L. D., Bursalioglu, A., Bühler, V., Çadek, M., Çetinçelik, M., Clay, G., Cortijos-Bernabeu, A., Damjanović, K., Dugue, T. M., Esberg, M., Esteban-Serna, C., . . . Folke, T. (2021). The general fault in our fault lines. Nature Human Behaviour, 5(10), 1369–1380. https://doi.org/10.1038/s41562-021-01092-x

Sah, P., Vilches, T. N., Moghadas, S. M., Fitzpatrick, M. C., Singer, B. H., Hotez, P. J., & Galvani, A. P. (2021). Accelerated vaccine rollout is imperative to mitigate highly transmissible COVID-19 variants. Eclinicalmedicine, 35, Article 100865. https://doi.org/10.1016/j.eclinm.2021.100865

Slade, M., & Priebre, S. (2001). Are randomised controlled trials the only gold that glitters. The British Journal of Psychiatry, 179, 286–287.

Stallard, N., Posch, M., Friede, T., Koenig, F., & Brannath, W. (2009). Optimal choice of the number of treatments to be included in a clinical trial. Statistics in Medicine, 28(9), 1321–1338.

Strengthening Democracy Challenge. (2021). Strengthening democracy challenge handbook. https://www.strengtheningdemocracychallenge.org/files/ugd/2107d4_a1bf6d473378c798e08b8ccad910d8ee.pdf
Tamir, M. (2009). What do people want to feel and why? Pleasure and utility in emotion regulation. *Current Directions in Psychological Science, 18*(2), 101–105. https://doi.org/10.1111/j.1467-8721.2009.01617.x

Tamir, M., Vishkin, A., & Gutentag, T. (2020). Emotion regulation is motivated. *Emotion, 20*, 115–119.

Tian, X., Li, C., Huang, A., Xia, S., Lu, S., Shi, Z., Lu, L., Jiang, S., Yang, Z., Wu, Y., & Ying, T. (2020). Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerging Microbes & Infections, 9*, 382–385. https://doi.org/10.1080/22221751.2020.1729069

Uhlmann, E. L., Ebersole, C. R., Chartier, C. R., Errington, T. M., Kidwell, M. C., Lai, C. K., McCarthy, R. J., Riegelman, A., Silberzahn, R., & Nosek, B. A. (2019). Scientific utopia III: Crowdsourcing science. *Perspectives on Psychological Science, 14*(5), 711–733. https://doi.org/10.1177/1745691619850561

Uschner, D., Hilgers, R. D., & Heussen, N. (2018). The impact of selection bias in randomized multi-arm parallel group clinical trials. *PLOS ONE, 13*(1), Article e0192065. https://doi.org/10.1371/journal.pone.0192065

Van Assche, J., Noor, M., Dierckx, K., Saleem, M., Bouchat, P., de Guissme, L., Bostyn, D., Carew, M., Ernst-Vintila, A., & Chao, M. M. (2020). Can psychological interventions improve intergroup attitudes post terror attacks? *Social Psychological and Personality Science, 11*, 1101–1109. https://doi.org/10.1177/1948550619896139

Van Bavel, J. J., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M. J., Crum, A. J., Douglas, K. M., Druckman, J. N., Drury, J., Dube, O., Ellemers, N., Finkel, E. J., Fowler, J. H., Gelfand, M., Han, S., Haslam, S. A., Jetten, J., & Willer, R. (2020). Using social and behavioral science to support COVID-19 pandemic response. *Nature Human Behaviour, 4*, 460–471. https://doi.org/10.1038/s41562-020-0884-z

Wason, J., Magirr, D., Law, M., & Jaki, T. (2016). Some recommendations for multi-arm multi-stage trials. *Statistical Methods in Medical Research, 25*, 716–727.

Wason, J. M. S., & Robertson, D. S. (2021). Controlling type I error rates in multi-arm clinical trials: A case for the false discovery rate. *Pharmaceutical Statistics, 20*(1), 109–116.

Wason, J. M. S., Stecher, L., & Mander, A. P. (2014). Correcting for multiple-testing in multi-arm trials: Is it necessary and is it done? *Trials, 15*, Article 364. https://doi.org/10.1186/1745-6215-15-364

Wilson, T. D., Aronson, E., & Carlsmith, K. (2010). The art of laboratory experimentation. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 51–81). John Wiley & Sons.

Wilson, T. D., & Juarez, L. P. (2015). Intuition is not evidence: Prescriptions for behavioral interventions from social psychology. *Behavioral Science & Policy, 1*, 13–20.

World Health Organization. (2020). *WHO R&D Blueprint: Novel coronavirus. An international randomised trial of candidate vaccines against COVID-19—Outline of solidarity vaccine trial*. https://cdn.who.int/media/docs/default-source/blue-print/who-covid-2019-solidarityvaccine-trial-expandedoutline-28may.pdf

World Health Organization. (n.d.). *WHO coronavirus (COVID-19) dashboard*. https://covid19.who.int/

Yokum, D., Lauffenburger, J. C., Ghazinouri, R., & Choudhry, N. K. (2018). Letters designed with behavioural science increase influenza vaccination in Medicare beneficiaries. *Nature Human Behaviour, 2*, 743–749.