a publication of the behavioral science & policy association



featured topic behavioral insights in the lab and around the world

behavioralpolicy.org

founding co-editors

Craig R. Fox (UCLA) Sim B Sitkin (Duke University)

senior policy editor

Varun Gauri (Princeton University)

bspa executive director

Kate B.B. Wessels

advisory board

Paul Brest (Stanford University) David Brooks (New York Times) John Seely Brown (Deloitte) Robert B. Cialdini (Arizona State University) Adam M. Grant (University of Pennsylvania) Daniel Kahneman (Princeton University) Faisal Naru (Policy Innovation Centre) Jeffrey Pfeffer (Stanford University) Denise M. Rousseau (Carnegie Mellon University) Paul Slovic (University of Oregon) Cass R. Sunstein (Harvard University) Richard H. Thaler (University of Chicago)

executive committee

Morela Hernandez (University of Virginia) Katherine L. Milkman (University of Pennsylvania) Daniel Oppenheimer (Carnegie Mellon University) Todd Rogers (Harvard University) David Schkade (UC San Diego) Joe Simmons (University of Pennsylvania)

bspa team

Kaye N. de Kruif, Managing Editor (Duke University) Carsten Erner, Statistical Consultant (FS Card) A. David Nussbaum, Director of Communications (Chicago) Ricki Rusting, Editorial Director Poruz Khambatta, Media Manager

consulting editors

Shlomo Benartzi (UCLA) Laura L. Carstensen (Stanford University) Susan T. Fiske (Princeton University) Carol L. Graham (Brookings Institution) Chip Heath (Stanford University) David I. Laibson (Harvard University) George Loewenstein (Carnegie Mellon University) Richard E. Nisbett (University of Michigan) M. Scott Poole (University of Illinois) Eldar Shafir (Princeton University)

policy editors

Henry J. Aaron (Brookings Institution) Matthew D. Adler (Duke University) Peter Cappelli (University of Pennsylvania) Thomas D'Aunno (NYU) J.R. DeShazo (UCLA) Brian Gill (Mathematica) Michal Grinstein-Weiss (Washington University) Ross A. Hammond (Brookings Institution) Ron Haskins (Brookings Institution) Arie Kapteyn (University of Southern California) John R. Kimberly (University of Pennsylvania) Mark Lubell (UC Davis) Annamaria Lusardi (George Washington University) Timothy H. Profeta (Duke University) Donald A. Redelmeier (University of Toronto) Rick K. Wilson (Rice University) Kathryn Zeiler (Boston University)

disciplinary editors

Behavioral Economics Senior Disciplinary Editor Associate Disciplinary Editors

Decision, Marketing, & Management Sciences

Senior Disciplinary Editor Associate Disciplinary Editors

> **Organizational Science** Senior Disciplinary Editors

Associate Disciplinary Editors

Social Psychology

Senior Disciplinary Editor Associate Disciplinary Editors

Senior Disciplinary Editor

Dean S. Karlan (Northwestern University) Oren Bar-Gill (Harvard University) Colin F. Camerer (California Institute of Technology) M. Keith Chen (UCLA) Julian Jamison (World Bank) Russell B. Korobkin (UCLA) Devin G. Pope (University of Chicago) Jonathan Zinman (Dartmouth College)

Eric J. Johnson (Columbia University) Linda C. Babcock (Carnegie Mellon University) Max H. Bazerman (Harvard University) Baruch Fischhoff (Carnegie Mellon University) John G. Lynch (University of Colorado) Ellen Peters (Ohio State University) John D. Sterman (MIT) George Wu (University of Chicago)

Carrie R. Leana (University of Pittsburgh) Jone L. Pearce (UC Irvine) Stephen R. Barley (Stanford University) Rebecca M. Henderson (Harvard University) Thomas A. Kochan (MIT) Christopher P. Long (St. John's University) Ellen E. Kossek (Purdue University) Elizabeth W. Morrison (NYU) William Ocasio (Northwestern University) Sara L. Rynes-Weller (University of Iowa)

Richard Larrick (Duke University) Dolores Albarracín (University of Illinois) Susan M. Andersen (NYU) Thomas N. Bradbury (UCLA) John F. Dovidio (Yale University) David A. Dunning (Cornell University) E. Tory Higgins (Columbia University) John M. Levine (University of Pittsburgh) Harry T. Reis (University of Rochester) Tom R. Tyler (Yale University)

David Grusky (Stanford University) Peter Hedstrom (Oxford University) Arne L. Kalleberg (University of North Carolina) James Moody (Duke University) Robert J. Sampson (Harvard University) Bruce Western (Harvard University)

Sociology

Associate Disciplinary Editors

Copyright © 2023 Behavioral Science & Policy Association SAGE Publications

ISSN 2379-4607 (print) ISSN 2379-4615 (online) ISBN 978-0-8157-3940-1 (pbk) ISBN 978-0-8157-3941-8 (epub) Behavioral Science & Policy is a publication of the Behavioral Science & Policy Association, P.O. Box 51336, Durham, NC 27717-1336, and is published twice yearly with SAGE, 2380 Conejo Spectrum Street, Thousand Oaks, CA 91320

Authorization to photocopy items for internal or personal use or the internal or personal use of specific clients is granted by SAGE Publications for libraries and other users registered with the Copyright Clearance Center Transactional Reporting Service, provided that the basic fee is paid to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For more information, please contact, CCC, at +1 978-750-8400 and online at www.copyright.com

This Authorization does not extend to other kinds of copying, such as copying for general distribution, or for creating new collective works, or for sale Specific written permission for other copying must be obtained from the Permissions Department, SAGE Publications.



İİ Editors' note

Features

1 Report

Saving for retirement: A real-world test of whether seeing photos of one's future self encourages contributions

Juan David Robalino, Alissa Fishbane, Daniel G. Goldstein, & Hal E. Hershfield

11 Review

Using communication to boost vaccination: Lessons for COVID-19 from evaluations of eight large-scale programs to promote routine vaccinations

Heather Barry Kappes, Mattie Toma, Rekha Balu, Russ Burnett, Nuole Chen, Rebecca Johnson, Jessica Leight, Saad B. Omer, Elana Safran, Mary Steffel, Kris-Stella Trump, David Yokum, & Pompa Debroy

45 Finding

Encouraging self-blinding in hiring Sean Fath, Richard P. Larrick, & Jack B. Soll

27 Finding

Can a visual values-affirmation intervention improve test scores of students in areas affected by crisis? Daniel D. Shephard, Ali Osseiran, & Fadi Makki

59

Review

Teamwork doesn't just happen: Policy recommendations from over half a century of team research Gudela Grote and Steve W. J. Kozlowski

78

Editorial policy

editors' note

Welcome back to *Behavioral Science & Policy* (*BSP*).

Perhaps you notice something different? With this issue, we are excited to begin a new publishing partnership with Sage Publications, which will enable us to bring the latest in actionable behavioral science to a broader audience and help us continue to enhance the impact of the work contained within these pages. The mission of *BSP* and its parent organization, the Behavioral Science & Policy Association, is to promote the thoughtful application of rigorous behavioral science to policy and practice in ways that serve the public interest. In short, we seek to bridge the divide between science and practice.

To do this, we publish articles that highlight novel and actionable insights from a wide range of social and behavioral sciences. Uniquely, *BSP* submissions are reviewed by both disciplinary scientists (to guarantee their rigor) and policy and practice analysts (to enhance their actionability). Moreover, authors of accepted manuscripts receive writing guidance from professional editors so that the articles are accessible to a diverse audience that includes scientists and policymakers in and out of government and members of the educated lay public.

The current issue features an excellent slate of articles that make good on this ambition. They provide important insights for readers interested in health care, financial decision-making, education, and management. Among the articles can be found empirical work and reviews, experiments that took place in the lab and in the field, and work conducted in the United States and abroad.

The first three articles all attempt to translate established laboratory insights into practices that can be carried out in more naturalistic environments.

Juan David Robalino, Alissa Fishbane, Daniel G. Goldstein, and Hal E. Hershfield provide a critical field test of a behavioral insight that was previously established only in laboratory settings. One of the obstacles to preparing for retirement is people's tendency to be biased toward spending money on things to be enjoyed in the present time, in part due to a lack of adequate consideration of the financial needs of their future selves. Laboratory experiments have found that providing people with age-advanced photographs of themselves can increase their stated motivation to save more for retirement—at least in hypothetical scenarios. This article presents a unique large-scale field test of this method among users of a Mexican mobile banking app, which yielded a statistically significant effect on real retirement saving decisions that was small but cost-effective.

Next, Heather Barry Kappes, Mattie Toma, Rekha Balu, Russ Burnett, Nuole Chen, Rebecca Johnson, Jessica Leight, Saad B. Omer, Elana Safran, Mary Steffel, Kris-Stella Trump, David Yokum, and Pompa Debroy take stock of a variety of vaccination-promoting communication strategies that were studied in eight largescale randomized controlled trials run by the U.S. Office of Evaluation Sciences prior the pandemic. These interventions are noteworthy for the scale of their application (having a median sample size of 55,000), their transparency (all trials were preregistered), and their focus on actual vaccination behaviors (rather than intentions to get vaccinated). Although most of these interventions were not successful, two proved somewhat effective in increasing vaccination rates: sending letters reminding older Medicare beneficiaries to be vaccinated for influenza and sending postcards reminding seniors in Louisiana to get a variety of vaccinations. The effects of these interventions were modest when compared with the greater effects of messaging on vaccination intentions seen in published academic literature, and so they serve as an important reality check about the perils of expecting stated intentions to translate into action outside the lab. The interventions also provide valuable insights for future researchers and practitioners attempting to promote vaccination using only messaging strategies.

Several previous studies have found that the test performance of members of stigmatized minority groups can be harmed by their worries about

confirming unflattering stereotypes about their group's academic abilities (for example, women being poor at math). Moreover, other studies have found that these effects can be buffered by providing people with an opportunity to engage in self-affirmation by writing about a value that is important to them (for example, friendships or religion). Daniel D. Shephard, Ali Osseiran, and Fadi Makki show that this technique can be extended to low-literacy populations (in this case, Syrian refugees in Lebanon) by using a values-affirming drawing exercise. In particular, students enrolled in an accelerated basic literacy and numeracy program who completed this intervention achieved higher subsequent test scores in Arabic than did students who did not complete the intervention (although scores on tests of English and mathematics were not significantly affected). This proof of concept will no doubt inspire follow-up attempts to expand the application of values-affirmation interventions around the world.

The final two articles in this issue explore the application of research insights to organizational practice.

Sean Fath, Richard P. Larrick, and Jack B. Soll present an intriguing study of participants who all had prior experience making hiring decisions. The researchers were interested in the conditions under which hiring managers choose to remain unaware of information about job candidates that is obviously biasing (for example, race or gender), biasing in less obvious ways (for example, a name or a photograph), or relevant (for example, college major or work experience). Participants chose to avoid biasing or potentially biasing information under three conditions: (a) when asked to indicate what applicant information they wanted to receive rather than what information they did not want to receive, (b) when making selections for other people rather than when making the selections for themselves, and (c) when the information was obviously biasing rather than when it was less obviously biasing. Although this study involves hypothetical decisions, the authors use their demonstration to provide practical advice for human resource officers. When explicit blinding policies are not feasible, human resource decision-makers can use these approaches to encourage hiring managers to voluntarily blind themselves from potentially biasing information.

In the final article in this issue, Gudela Grote and Steve W. J. Kozlowski synthesize many decades of research on teamwork, a critical capability for managing crisis and fostering innovation. Fortunately, the capacity for better teamwork can be achieved through training and effort. The authors distill policy recommendations for educators, regulators, and organizational leaders.

In the eight years since the launch of BSP, we have been excited to witness an explosion of interest in applied behavioral science among policymakers and other practitioners. Meanwhile, we have been gratified to see a number of disciplinary journals publish more applied work and field studies, and we have noted that the list of new translational journals is growing. BSP is a unique outlet for authors who wish to see their best behavioral science research and insights have impact both inside and outside academia and who, by working a little harder (with our help), are able to translate their findings into immediately actionable recommendations expressed in language that is accessible to a broad audience. In this way, we hope that our journal will continue to serve an important role in bridging the divide between science and practice.

Craig R. Fox & Sim B Sitkin Founding Co-Editors





Saving for retirement: A real-world test of whether seeing photos of one's future self encourages contributions

Juan David Robalino, Alissa Fishbane, Daniel G. Goldstein, & Hal E. Hershfield

abstract

One psychological barrier to putting money aside for retirement may be an inability to fully empathize with the economic woes of one's future self. In tests of ways to lower this barrier, previous studies have had experimental participants interact with visualizations of their future selves. Despite the promise shown by such interventions in small-scale tests in the lab, little is known about their effectiveness in the real world. Our research evaluates the effectiveness of an aging filter (that is, software that creates an image of how a participant might look when older) in a randomized field study involving nearly 50,000 people saving for retirement in Mexico. The intervention, carried out over a month, modestly increased the number of account holders who made one-time contributions (from 1.5% in the control group to 1.7% in the treatment group, representing a 16% increase), as well as the value of those contributions. Although the total amount of money put aside was modest and the number of sign-ups for a recurring contribution savings program did not change significantly, this intervention proved cost-effective: It increased savings at a rate almost 500 times the cost of the intervention. Such psychologically informed interventions can effectively complement other initiatives to encourage people to save for retirement.

Robalino, J. D., Fishbane, A., Goldstein, D. G., & Hershfield, H. E. (2023). Saving for retirement: A real-world test of whether seeing photos of one's future self encourages contributions. *Behavioral Science & Policy*, 9(1), 1–9. https://doi.org/10.1353/xxxxxxx

Note that the set of t

In many nations, retirement savings fall far below the amount considered sufficient to maintain preretirement standards of living. In Mexico, the retirement pension system changed in 1997 to a defined contribution system, in which employers are required to deposit a set amount of each paycheck into a retirement account, with employees then encouraged to make additional contributions themselves. The mandatory contributions are set by the government; as we write this article, the rate is just 6.5% of a worker's salary. Through these mandatory contributions alone, workers are projected to receive less than 30% of their preretirement salary in retirement, far less than the traditional benchmark of 75% that is generally considered sufficient to provide a secure retirement.

More than 99.5% of Mexico's 40.5 million registered account holders do not make any additional contributions in a given year.¹ Moreover, 60% of workers are in the informal work sector and do not have any mandatory retirement savings, making their financial futures particularly vulnerable.²

Countries around the world take various approaches to encourage retirement saving. Some, for instance, automatically enroll workers in savings plans and require them to actively opt out if they do not want to participate. Yet even when such choice architecture interventions are used, there still may be room to increase both participation in plans and the amount saved. In Mexico, as in many other nations, increasing retirement contributions could improve the financial situations of millions. Although some people may not save for retirement because they are struggling to pay for the essentials of day-to-day life, many are also deterred by psychological barriers. For instance, some may lack the willpower to save when facing the temptation to spend on something they want now,^{3,4} or they may find it hard to predict how they will feel in the future about decisions they make today.⁵ In this article, we focus on a related barrier: the difficulty people have identifying with—and feeling concern for their future selves.

Previous theoretical work suggests that people may think about their future selves—the selves who may benefit from or be punished by choices made today—as if they are other people altogether.^{6,7} Thinking of the future self as another person can, in theory, have a detrimental effect on long-term decision-making and planning. For example, it is easy to eat too much pizza and chips in the present moment or splurge on an unaffordable car when the future consequences of obesity or debt are thought of as "someone else's problem." In this light, impatience is simply a form of self-interested behavior—a person acts for the benefit of their current, present self rather than their future self.

Of course, people do at times make sacrifices for others, particularly people they feel close to, like family and friends. Therefore, it might be the case that people will make sacrifices for their future selves if they feel as connected to their future selves as they do to their friends and family.⁸ This theory has been borne out by some early, small-scale studies in the lab. In one study, for example, people who felt more connected to their future selves accrued more assets in savings, even when the researchers controlled for demographic factors.⁹ And in another study, high school students who felt more connected to their future selves maintained higher levels of academic performance.¹⁰

How can the gap between current and future selves be narrowed? Consider a technique used by charitable organizations: Presenting people with salient and vivid representations of charity recipients (for example, through stories about who the recipients are and how the charity has helped them) makes them more likely than those who are not presented with such representations to make donations.¹¹

Drawing on that finding, researchers have attempted to increase the vividness of future selves to induce people to take better care of those distant selves. For instance, when participants were shown vivid, visual representations of their future selves in laboratory contexts, they were more likely to make contributions to a hypothetical long-term savings account¹² and opt for more ethical paths when given the opportunity to cheat.¹³

Despite the promise seen in the results of such interventions from small studies conducted in lab settings, little is known about their effectiveness in large-scale field settings with real consequences. This is an important gap in the research: Field trials can help to inform the theories that motivate an intervention,¹⁴ and achieving success in the field is critical before policymakers can justify widely deploying any intervention to address societal problems, such as inadequate saving for retirement.¹⁵ This is the gap we set out to address in our study.

Method

We ran our field experiment in Mexico throughout August 2018 on all 48,853 clients who had retirement accounts with a specific fund administrator and used an online mobile banking app called AforeMóvil to set up and make voluntary contributions to those accounts. This popular app is provided by the government regulatory commission CONSAR (Comisión Nacional del Sistema de Ahorro para el Retiro) and is available to all retirement savings plan providers (known in Mexico as *Administradoras de Fondos para el Retiro*). We split the group in half randomly on July 25, 2018, allocating 24,427 account holders to the treatment group and 24,426 to the control group.

The two groups did not differ in terms of their members' gender, initial savings, and age (see Table S1 in the Supplemental Material for a detailed analysis). Seventy-three percent of the account holders from the whole study group were men, and the average total retirement savings balance was 212,200 Mexican pesos (approximately US\$10,500), out of which 8,650 pesos (approximately US\$420) were voluntary savings. For context, the average monthly household income per capita in Mexico at the time was approximately US\$240.¹⁶ Sixty-one percent of account holders were in the highrisk investment portfolio designed for those 36 years of age and younger, 23% were in the portfolio designed for 37- to 45-year-olds, 15% were in the portfolio for 46- to 59-year-olds, and 1% were in the portfolio for those over 60.

The intervention for the treatment group involved three main stages. First, account holders received an invitation message with a link to meet their future self. Second, those clicking the link were directed to a web page where they could take a selfie and see an age-progressed rendering of their future self (see Figure 1). A computer algorithm, made by an anonymous commercial company in collaboration with our research group, automatically aged the person's image to approximate what they would look like in about 20 years. Third, after taking this step, account holders saw a message below their aged photo asking how much they would like to save for the person in the picture to live well, along with a link to the savings page on the app. Account holders could then choose to enroll in the recurring-deposit program, make a one-time contribution, or do both.

To increase the likelihood of attracting account holders' attention, we planned to send each account holder nine invitation messages-three emails, three mobile phone text messages, and three push notifications in the app-over the course of one month. We also used three different themes along the following lines for the phrasing in the invitations: "How will you look in old age?" (aiming to spark a sense of curiosity); "message from your 'future self'" (aiming to grab attention through mystery); and "limited time, try it today" (drawing on a sense of urgency).¹⁷ (See Figures S1-S7 in the Supplemental Material for the full text of the email, text, and app messages in Spanish and English.) The order of the type of communication used (email, text, or push notification) and the message theme was randomized. Whether or not account holders

11:15 AM AT&T 穼 11:16 AM AT&T 🗢 \$ 94% 🔳 \$ 93% + \sim changemyface.com C changemyface.com C sura AFORE Adelanta el tiempo conoce tu versión 2050 Hola Juan David, ate que Juan Da móvil co filtro tercera edad de fotos se Haz clic aquí Sonríe, tómate una selfie y en ¿Cuánto quieres guardar para pocos segundos te presentáremos a tu "yo futuro" tu "Yo futuro" no quiere que sigas es que él viva bien? liza este correo desde tu dispositivo móvil. Siguenos en: y f O Toma una selfie Û ſŊ SURA te invita

Figure 1. The results of the aging filter on a selfie

Note. The far left panel shows a message inviting account holders to meet their future selves; the English and Spanish versions of all invitations may be seen in the figures in the Supplemental Material. The translation of the text from the middle image reads, "Smile, take a selfie and in a few seconds we will introduce you to 'your future self." The translation of the text from the right image reads, "How much do you want to save so that he lives well?"

clicked on the links, took a selfie, or made a contribution, they received all scheduled invitations throughout the month. Because of a logistical issue, one of the planned emails was, in the end, not sent, so each account holder received a total of eight invitations.

The control group also received messages (by email, text, and push notification, at the same times as the treatment group) encouraging them to save. These messages had themes similar to those in the messages sent to the treatment group but did not mention the aging photo filter. In translation, the text messages, for example, read, "Do you want to save for retirement? Click here!"; "Have you started saving for retirement? Click here and program your savings!"; and "Don't waste more time, click here and save TODAY for a better future." Each control message had a link directly to the saving app.

Results

Almost half of the treatment group chose to click through to the photo filter page at least once. A total of 11,092 account holders (45% of the treatment group) at some point clicked through, together making 32,615 visits to the web page. Of those who visited the photo filter web page, more than half took selfies: 6,843 account holders (28% of the treatment group) took a total of 13,041 selfies. Then 2,585 account holders (10.5% of the treatment group) clicked on the AforeMóvil contribution link a total of 3,013 times (see Table 1).

In terms of engagement, email and mobile phone text messages proved more effective than the app's push notifications at generating initial interest: 42% of first visits to the photo filter web page came from emails, 40% from text messages, and 18% from push notifications. The theme of the message also mattered: 44% of first visits came from the "how will you look in old age" communications, 39% came from the "message from your 'future self'" communications, and 17% came from the limited-time communications. When the effects of all communications were analyzed, we found that the same patterns held, and the differences were significant at the p < .01 level. (See Table S6 in the Supplemental Material for more details, and see note A for a discussion of the statistical terms used in this article.)

Table 1.	Summary of	actions	by account	holders in	the study
----------	------------	---------	------------	------------	-----------

	Control group (<i>n</i> = 24,426)		Treatment group (<i>n</i> = 24,427)	
Action	n	%	n	%
Visited the selfie website			11,092	45
Took selfies			6,843	28
Clicked from selfies to contributions in the AforeMóvil app			2,585	10.5
Made a one-time contribution	355	1.5	412	1.7

Note. The study was conducted in August 2018. The treatment led to a statistically significant increase of 0.2 percentage point in the number of account holders who made a one-time contribution (p < .05), although the number of people who made contributions was very small in both groups.

Overall, we found that men were more likely than women to take a selfie, click through to the app, and make a contribution. Also, the probability increased with the age, rising from one age group to the next. Looking at retirement account balances and salary levels, we found that the probability of completing the entire process increased with the account balance of the account holder and decreased with the salary level. (Overall, these findings were significant at p < .05; see Table S2 in the Supplemental Material for a detailed analysis.)

As we noted earlier, account holders could enroll in the recurring-deposit program, make a one-time contribution, or do both. Before starting our study, we preregistered (https:// aspredicted.org/blind.php?x=uu88y4) two main outcomes-whether clients enrolled in a recurring-deposit program and whether clients made any voluntary contribution (that is, whether they made a one-time contribution or signed up for the recurring-deposit program)—as well as the magnitude of the contributions. Given the extremely low retirement savings contribution rate in Mexico, upon reflection, we realized we had failed to include a simple analysis of one-time contributions. Because this information is of practical interest, we report three outcomes: recurringdeposit sign-ups, one-time contributions, and whether account holders made any voluntary contribution.

We found no significant effects in the percentage of account holders who enrolled in the recurring-deposit program, perhaps because signing up for recurring deposits is a large commitment and thus hard to influence, nor did we find a statistically significant effect on the likelihood of making any voluntary contribution. Nonetheless, the treatment modestly increased the number of account holders who made one-time contributions to a statistically significant extent: 1.5% of the control group made contributions (355 people), compared with 1.7% of the treatment group (412 people; p < .05). (See Table 1; also see Table S3 in the Supplemental Material for a detailed analysis.) Although a 0.2% increase sounds small, this means that 16% more people in the treatment group than in the control group made contributions.

Given that over 98% of account holders in the study made no contributions to their retirement savings, any effects among the people who did contribute get diluted in simple statistical regressions, the kinds of analyses often done with this kind of data. A more informative approach is to use Tobit regression-a statistical approach that isolates the effect from the subset of people making nonzero contributions. This is the common model used to analyze savings when many accounts have a zero balance. With this approach, we found that the intervention greatly raised the average amount of single contributions (conditional on a contribution being made): The average contribution in the control group was 3,063 pesos, and the average contribution in the intervention group was 1,327 pesos higher (a 43% increase). (See Table 2; see also Table S4.) Not all people in the intervention group completed the process (in other words, not everyone who had the opportunity to take a selfie did so); this

Group	Any voluntary contribution	Recurrent contributions	One-time contribution
Treatment	52.85	-18.83	1,327.48**
	(92.03)	(46.12)	(592.03)

Table 2. Results: The peso amounts contributed to savings by the treatment group relative to control group, conditional on making a contribution

Note. We used a Tobit regression to see the effect of the intervention on the peso amount of retirement contributions among those who made a contribution. This approach is the one commonly used to analyze savings when many savings accounts have zero balance. Standard errors are in parentheses. **p < .05.

suggest that the effect we found is a lower bound of the effect of the intervention on the people who went through the intervention.

Collectively, account holders in the treatment group contributed 54% more to retirement in August than the control group did: 1,675,974 and 1,087,422 pesos, respectively. Although these seem like large increases in money, these results should be treated with caution given the very small number of people making the contributions leading to these amounts.

Finally, we explored whether income level, age, gender, and formal employment influenced the retirement contributions made. (We used mandatory contributions as a proxy for formal employment, given that these contributions are made by employers.) We found some evidence that the intervention increased recurring savings of individuals aged 37-45 years by 305 pesos compared with the recurring savings of individuals in the treatment group who were younger than 36 years or older than 46 years (p < .05). We also found that for the people in the treatment group who made voluntary contributions, those contributions increased by about 16% per additional peso of income (p < .01; see Table S5 in the Supplemental Material for detailed analysis). In other words, those with a higher income may have been more influenced by this intervention. This pattern may reflect the fact that for many workers with limited incomes, saving even a little is a considerable burden, and this intervention may not influence such workers.

These results should be interpreted with care, however, given the large number of tested parameters in this model and the noisy nature of our income measure. Indeed, employers in Mexico widely underreport income because of tax incentives and to reduce the money they must contribute to retirement accounts.

Perhaps the most interesting aspect of the intervention is its cost-effectiveness. The development of the aging photo filter itself and the host web page cost 139,490 pesos (US\$7,000). The text messages cost 0.56 peso (US\$0.028) per message, in this case totaling roughly 82,940 pesos (US\$4,130), and push notifications and emails had no direct costs outside of the time and effort required to administer them.

At present, about 20 million people in Mexico have retirement savings plans, 2.2 million of whom use AforeMóvil. Using emails and push notifications alone to share the aging photo filter web page, it would be possible to scale this intervention to all 2.2 million current AforeMóvil users for about 112,785 pesos (US\$5,500) per month, given the costs of the photo filter license and AforeMóvil staff time.

Extrapolating from our results that the intervention increased the number of contributors by 0.2%, we calculate that this expansion could result in an additional 4,400 contributors out of 2.2 million people, generating an increase of 53,004,600 pesos (US\$2,584,782) in voluntary retirement savings in just one month. That translates to about 470 pesos in savings generated per 1 peso invested.

Discussion

Around the world, policymakers are keen to encourage people to save more for their retirement. To help, researchers have introduced a variety of interventions based on principles of social psychology and behavioral economics. For example, requiring employees to opt out of a default defined-contribution plan rather than waiting for them to opt in to such a program can increase the number of people who participate in a retirement plan. Such strategies have been shown to greatly increase the number of employees who contribute to their retirement funds.¹⁸ Similarly, introducing an automatic escalation feature for retirement saving plans (in which workers have their paycheck contributions automatically go up each fiscal year without having to take any action) has also had a positive effect on saving rates.¹⁹

But these effective interventions are not used everywhere. Mexico, for one, does not use them. Even in countries where they are used, employers must implement them. Also, workers across the world are increasingly turning to contract work: Some estimates predict that close to 50% of the American workforce could be contract workers by 2030.²⁰ So it would be useful to come up with additional ways to encourage individuals to increase their voluntary retirement fund contributions.

The treatment we implemented—exposure to age-progressed images—led to a 16% increase in the number of account holders who made one-time contributions to their retirement savings accounts (off a modest baseline) as well as an increase in the average amount they saved. Our results suggest that investing in the intervention on a large scale would be cost-effective: When scaled up across the 2.2 million account holders in Mexico, the approach could result in an additional 470 pesos saved per 1 peso spent on the intervention.

Given the complex set of factors that affect who saves what for their retirement, it is reasonable to expect that the effects of any one intervention should be small.²¹

Our study did have some limitations. Because we could test this intervention only with the customers who already had the online banking app, our results may be less relevant for people who do not have the app or even a bank account, which may be true more often for people with low incomes and informal or no employment. Further, because the aging filter was hosted on an external website and not directly within AforeMóvil, some account holders may have had logistical difficulties going through the steps of the whole intervention process, from taking a photo to making a retirement saving deposit. All of the results should also be treated with some caution given the large number of outcomes we analyzed, which may have inflated the chance of finding significant effects.

The design of this field study does not allow us to determine why viewing age-progressed images led people to contribute more money to retirement. In theory, such images should enhance an emotional connection to and concern for the well-being of one's future self, as previous research has found.^{12,22} However, other factors could help to explain the increase. For example, the treatment may have led individuals to simply pay more attention to the savings decision, and this additional attention could have increased their contributions. If possible, future research—in the field and in the lab—should examine such alternatives.

Our study answers a growing call for psychological research to address pressing societal concerns.²³ We have shown that helping people to construct vivid, realistic images of their future selves can affect their real decisions to increase savings for the future. However this effect is accomplished, the findings suggest that from a practical standpoint, using age-progressed photos could complement existing successful strategies for increasing retirement savings.

endnote

A. Editors' note to nonscientists: For any given data set, the statistical test used-such as the chi-square (χ^2) test, the *t* test, or the *F* test–depends on the number of data points and the kinds of variables being considered, such as proportions or means. The p value of a statistical test is the probability of obtaining a result equal to or more extreme than would be observed merely by chance, assuming there are no true differences between the groups under study (this assumption is referred to as the null hypothesis). Researchers traditionally view p < .05 as the threshold of statistical significance (also referred to as a 5% significance level), with lower p values indicating a stronger basis for rejecting the null hypothesis. Standard deviation is a measure of the amount of variation in a set of values. Approximately two-thirds of the observations fall between one standard deviation below the mean and one standard deviation above the mean. Standard error uses standard deviation to determine how precisely one has estimated a true population value from a sample. For instance, if one took enough samples from a population, the sample mean ±1 standard error would contain the true population mean around two-thirds of the time.

author affiliation

Robalino: Universidad San Francisco de Quito. Fishbane: ideas42. Goldstein: Microsoft Research. Hershfield: University of California, Los Angeles. Corresponding author's e-mail: hal.hershfield@anderson.ucla.edu.

author note

The authors acknowledge generous funding from MetLife Foundation and assistance with intervention design and implementation from Andrew Fertig and Marcela Cheng Oviedo.

supplemental material

- http://behavioralpolicy.org/journal
- Additional Tables & Figures

references

- 1. Comisión Nacional del Sistema de Ahorro para el Retiro. (2018). Informe trimestral al H. Congreso de la Unión sobre la situación del SAR [Quarterly report to the H. Congress of the Union on the situation of the SAR]. Gobierno de México. https://www.gob.mx/cms/ uploads/attachment/file/323965/ Informe_Trimestral_1T2018.pdf
- 2. Organisation for Economic Cooperation and Development. (2016). OECD reviews of pension systems: Mexico. https://doi. org/10.1787/9789264245938-en
- Carlson, S. M., Shoda, Y., Ayduk, O., Aber, L., Schaefer, C., Sethi, A., Wilson, N., Peake, P. K., & Mischel, W. (2018). Cohort effects in children's delay of gratification. *Developmental Psychology*, 54(8), 1395–1407. https:// doi.org/10.1037/dev0000533
- Watts, T. W., Duncan, G. J., & Quan, H. (2018). Revisiting the marshmallow test: A conceptual replication investigating links between early delay of gratification and later outcomes. *Psychological Science*, 29(7), 1159–1177. https://doi. org/10.1177/0956797618761661
- Wilson, T. D., & Gilbert, D. T. (2005). Affective forecasting: Knowing what to want. Current Directions in Psychological Science, 14(3), 131–134. https://doi. org/10.1111/j.0963-7214.2005.00355.x
- Hershfield, H. E. & Bartels, D. (2018). The future self. In G. Oettingen, A. T. Sevincer, & P. M. Gollwitzer (Eds.), *The psychology of thinking about the future* (pp. 89–109). Guilford Press.
- 7. Parfit, D. (1984). *Reasons and persons*. Oxford University Press.
- 8. Whiting, J. (1986). Friends and future selves. *The Philosophical Review*, 95(4), 547–580.
- Ersner-Hershfield, H., Garton, M. T., Ballard, K., Samanez-Larkin, G. R., & Knutson, B. (2009). Don't stop thinking about tomorrow: Individual differences in future self-continuity account for saving. *Judgment and Decision Making*, 4(4), 280–286. http:// decisionsciencenews.com/sjdm/journal. sjdm.org/9310/jdm9310.pdf
- Adelman, R. M., Herrmann, S. D., Bodford, J. E., Barbour, J. E., Graudejus, O., Okun, M. A., & Kwan, V. S. Y. (2017). Feeling closer to the future self and doing better: Temporal psychological mechanisms underlying academic performance. *Journal of Personality*, *85*(3), 398–408. https://doi.org/10.1111/ jopy.12248

- Small, D. A., & Loewenstein,
 G. (2003). Helping a victim or helping the victim: Altruism and identifiability. Journal of Risk and Uncertainty, 26(1), 5–16. https://doi. org/10.1023/A:1022299422219
- Hershfield, H. E., Goldstein, D. G., Sharpe, W. F., Fox, J., Yeykelis, L., Carstensen, L. L., & Bailenson, J. N. (2011). Increasing saving behavior through age-progressed renderings of the future self. *Journal of Marketing Research*, 48(SPL), S23–S37. https://doi. org/10.1509/jmkr.48.SPL.S23
- Van Gelder, J.-L., Luciano, E. C., Kranenbarg, M. W., & Hershfield, H.
 E. (2015). Friends with my future self: A longitudinal vividness intervention reduces delinquency. *Criminology*, 53(2), 158–179. https://doi. org/10.1111/1745-9125.12064
- 14. Walton, G. M., & Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. *Psychological Review*, 125(5), 617–655. https://doi.org/10.1037/ rev0000115
- Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., Tucker-Ray, W., Congdon, W. J., & Galing, S. (2017). Should governments invest more in nudging? *Psychological Science*, 28(8), 1041–1055. https://doi. org/10.1177/0956797617702501
- 16. Instituto Nacional de Estadística, Geografía e Informática. (2018). Encuesta Nacional de Ingresos y Gastos de los Hogares. ENIGH 2018 nueva serie [National Survey of Household Income and Expenses. ENIGH 2018 new series]. https://www.inegi.org.mx/programas/ enigh/nc/2018/#Tabulados
- 17. Cialdini, R. B. (1984). *Influence: The psychology of persuasion*. William Morrow.
- Madrian, B. C., & Shea, D. F. (2001). The power of suggestion: Inertia in 401(k) participation and savings behavior. *The Quarterly Journal of Economics*, 116(4), 1149–1187. https://doi. org/10.1162/003355301753265543
- Benartzi, S., & Thaler, R. H. (2013, March 8). Behavioral economics and the retirement savings crisis. *Science*, 339(6124), 1152–1153. https://doi. org/10.1126/science.1231320
- Government Accounting Office. (2015). Contingent workforce: Size, characteristics, earnings, and benefits (GAO-15-168R). https://www.gao.gov/ products/gao-15-168r

- Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2022). Small effects: The indispensable foundation for a cumulative psychological science. *Perspectives on Psychological Science*, 17(1), 205–215. https://doi. org/10.1177/1745691620984483
- 22. Ganschow, B., Cornet, L., Zebel, S., & van Gelder, J.-L. (2021). Looking back from the future: Perspective taking in virtual reality increases future selfcontinuity. *Frontiers in Psychology*, *12*, Article 664687. https://doi.org/10.3389/ fpsyg.2021.664687
- Berkman, E. T., & Wilson, S. M. (2021). So useful as a good theory? The practicality crisis in (social) psychological theory. *Perspectives on Psychological Science*, 16(4), 864–874. https://doi. org/10.1177/1745691620969650



Get Immunized Today*

Vacúnate hoy mismo*

- Flu (Gripe)
- COVID-19
- Shingles (Herpes zóster)
- Pneumonia (Neumonía)
- Tetanus (Tétanos)
- Hepatitis
- HPV (VPH)

Walk-ins welcome Bienvenida sin cita

Subject to availability and applicable state law. Age restrictions apply. Ask your pharmacist for details.

*La mayoria de los planes de seguro cubren por completo muchas vacunas. Sujeto a disponibilidad y leyes estatales aplicables. Se aplican restricciones de edad. Pregunta a tu farmacéutico.





review

Using communication to boost vaccination: Lessons for COVID-19 from evaluations of eight largescale programs to promote routine vaccinations

Heather Barry Kappes, Mattie Toma, Rekha Balu, Russ Burnett, Nuole Chen, Rebecca Johnson, Jessica Leight, Saad B. Omer, Elana Safran, Mary Steffel, Kris-Stella Trump, David Yokum, & Pompa Debroy

abstract

The COVID-19 pandemic has added new urgency to the guestion of how best to motivate people to get needed vaccines. In this article, we present lessons gleaned from government evaluations of eight large randomized controlled trials of interventions that used direct communications to increase the uptake of routine vaccines. These evaluations, conducted by the U.S. General Services Administration's Office of Evaluation Sciences (OES) before the start of the pandemic, had a median sample size of 55,000. Participating organizations deployed a variety of behaviorally informed direct communications and used administrative data to measure whether people who received the communications got vaccinated or took steps toward vaccination. The results of six of the eight evaluations were not statistically significant, and a meta-analysis suggests that changes in vaccination rates ranged from -0.004 to 0.394 percentage points. The remaining two evaluations yielded increases in vaccination rates that were statistically significant, albeit modest: 0.59 and 0.16 percentage points. Agencies looking for cost-effective ways to use communications to boost vaccine uptake in the field-whether for COVID-19 or for other diseasesmay want to evaluate program effectiveness early on so messages and methods may be adjusted as needed, and they should expect effects to be smaller than those seen in academic studies.

Kappes, H. B., Toma, M., Balu, R., Burnett, R., Chen, N., Johnson, R., Leight, J., Omer, S. B., Safran, E., Steffel, M., Trump, K.-S., Yokum, D., & Debroy, P. (2023). Using communication to boost vaccination: Lessons for COVID-19 from evaluations of eight large-scale programs to promote routine vaccinations. *Behavioral Science & Policy*, *9*(1), 11–24. https://doi.org/10.1353/xxxxxxxx

ver since vaccines for COVID-19 became available, public health officials have tried many strategies to induce as many people as possible to roll up their sleeves.¹ Yet, at the time of this writing, participation in vaccine programs has been disappointing. Rates of uptake for many vaccines fall well below public health recommendations, both in the United States^{2,3} and in other countries.^{4,5} In the United States, uptake of COVID-19 vaccinations has also been lower than expected.

Direct communication to individuals is a commonly used, relatively inexpensive tool for trying to increase vaccination rates, and communication "to enhance informed vaccine decision-making" is one of the five goals of the U.S. National Vaccine Plan.² The approach makes sense: Communications have the potential to address a number of behavioral barriers to vaccination. Individuals may be unaware that a vaccine is available and recommended for them, may not believe that a particular vaccination is safe or effective, may not form an intention to get vaccinated, or may not remember or be able to act on an intention to vaccinate. Research in behavioral science provides insight on how to design letters, emails, and other direct communications to overcome such barriers.6-8 For example, research suggests that particular kinds of messages have the potential to influence behavior, such as those that tap into people's natural aversion to risk, provide the perspective of a hypothetical individual facing a decision, or reinforce good decision-making by emphasizing that a desired action is the norm.

Nevertheless, just how large a difference government communications can make has been unclear. In this article, we discuss a set of studies that presented an unusual opportunity to evaluate such interventions in a large-scale, real-world context. An analysis of this work offers lessons that might guide the use and evaluation of communications designed to improve uptake of vaccines against COVID-19 and other infectious disease.

The Evaluations in Detail

The research we review in this article was conducted by the U.S. General Services

Administration's Office of Evaluation Sciences (OES), a team of interdisciplinary experts who work across the U.S. government to help agencies build and use evidence, including behavioral insights, for the public good. Between 2015 and 2019, OES designed and tested an array of direct communications about vaccination in eight large-scale, randomized controlled trials-gold-standard experiments in which participants are assigned randomly to treatment and control groups to limit bias and enable researchers to explore cause-and-effect relationships. OES conducted the evaluations (known as the OES vaccination portfolio) in collaboration with a private health facility, a city department of health, a state department of health, three Veterans Health Administration health care systems, and one division of the U.S. Department of Health and Human Services.

The evaluations had a median sample size of 55,000, which is considerably larger than that reported in most behavioral science studies, as well as other appealing features. The interventions aimed to increase vaccination rates in populations that public health experts had strongly recommended be vaccinated, such as young children, pregnant women, and older adults. Several samples had high proportions of individuals from groups that have historically had lower vaccination rates. More than half the patients included in one of the evaluation's samples at a Veterans Affairs facility, for example, were African American. The interventions were wide-ranging. Some experiments used email, postcard, letter, or social media notifications to convey messages to potential vaccine recipients. Others used very different strategies: In one, school administrators received a formal report card of a school's vaccination compliance rate, and in another, clinicians received reminders through a hospital's electronic health record (EHR) system. The behavioral insights that informed the interventions also varied. Behavioral studies have tested strategies such as reminders, prompts that encourage recipients to make a plan to get vaccinated, messages that emphasize social norms, communications designed to be persuasive, and variations in the source and timing of messages. All these strategies were used in one or more of the interventions OES evaluated.

Although the OES evaluations focused on routine vaccinations, the findings are relevant to addressing the ongoing challenge of COVID-19 in part because, as is true for routine vaccinations, the challenge of achieving and maintaining widespread immunization is expected to continue. Many of the OES evaluations were implemented in the midst of broader vaccination campaigns, which will also be necessary to continue to fight COVID-19.

We selected the OES vaccination portfolio for analysis for another reason as well: These evaluations overcome some drawbacks of many other investigations into the effects of communications designed to influence behavior. Although the amount of research on using communications to alter behavior has increased rapidly and some published experiments show measurable impacts, some of these effects have been hard to replicate in the real world. A recent analysis of the literature on the use of nudges helps to explain why. Nudges, which often take the form of communications to influence behavior, are light-touch interventions that aim to alter people's behavior without constraining choice or providing significant economic incentives. Journal articles reporting on academic studies of nudges show effects that are 7.3 percentage points higher, on average, than those seen in evaluations conducted by government units. The analysis suggests that a combination of publication bias and low statistical power can account for the gap.9 Publication bias is the tendency to publish only statistically significant results. Such selective publication of results has been found to inflate expectations of actual effects and boost the likelihood of false positive findings.¹⁰ Statistical power is a study's ability to detect an effect if there is one. In general, published studies on communication interventions have had small sample sizes, which

Figure 1. Overview of Office of Evaluation Sciences vaccination uptake evaluations showing population segments that were sampled, intermediaries in the communication chain, sample sizes, & the modes of communication



Note. EHR = electronic health record. Ages are given in years. Evaluation details are in Table 1.

limit their power and the strength of the conclusions that can be drawn from them.

In the case of the vaccination portfolio, OES provided detailed preanalysis plans and committed to sharing the results of all evaluations; it has no "file drawer" where results are stashed away if they are not significant. The results of every evaluation of communications encouraging vaccination uptake conducted by OES from 2015 through 2019 have been reported on the OES website, and all evaluations are summarized here to avoid publication bias.

These evaluations were carefully designed to have high statistical power so as to detect even tiny effects. *Minimum detectable effect*, or MDE, is a measure of the sensitivity of a study; it is the smallest effect that, if it exists, would have an 80% chance of being detected. The OES evaluations had MDEs as small as 0.04 percentage points, and all but one had an MDE smaller than 1.7 percentage points. This made the evaluations powerful enough to detect the effects in the range of 2 to 4 percentage points that had been reported by two similar, related studies.^{11,12}

Results of the OES Evaluations

Table 1 contains a summary of the results of the eight evaluations. The communications used in each study can be seen by visiting https:// oes.gsa.gov/vaccines/ and clicking on the View Vaccination Portfolio Intervention Pack (PDF) button. Briefly, the interventions and results were as follows, listed roughly in the order in which they were done.

Evaluation 1

Letters encouraging flu vaccination were sent to Medicare beneficiaries in the experimental groups of this study, which was conducted in 2014–2015. A total of 227,955 beneficiaries received either no letter (the control group) or one of four versions of a letter encouraging vaccination (the experimental groups). The versions incorporated language that drew on past behavioral research. Study participants who received a letter were more likely to get their shot, although the version received made no difference.^{13,14}

Evaluation 2

Messages encouraging flu vaccination were sent to a randomly selected subset of 2,002 pregnant women through a Duke University Health System EHR messaging system in this study, conducted in 2016–2017. The messages noted that pregnant women are at greater risk of contracting the flu and that the vaccine provides protection for both mother and infant. The messages reminded patients that they could receive the vaccine at their next scheduled obstetric appointment. The rates of vaccination did not differ significantly between women who received the messages and women who did not.¹⁵

Evaluation 3

Varied social media advertisements promoted influenza and whooping cough vaccination for potentially pregnant women in this study, conducted in 2017. This campaign reached 591,221 women ages 20–34 years. It did not measure vaccination rates but instead analyzed click-through rates for four different messages to determine which messages motivated viewers to seek more information. The study found no statistically significant differences in the responses to the ads.¹⁶

Evaluation 4

In this study, conducted in 2017–2018, the Louisiana Department of Health sent postcard reminders to 208,867 residents ages 65–70 years who were overdue for any of four vaccines. Postcards were sent on a staggered schedule over a season, enabling timing to be used to create treatment and control groups. A reminder sent in October had a small but statistically significant effect on vaccine uptake. Two rounds of postcards sent to different groups in November and December had no effect.^{17,18}

Evaluation 5

In a study conducted in 2017–2018, the health department of a midsized city with 700 schools and daycare centers sent to randomly selected school leaders an immunization report card highlighting their school's immunization compliance in comparison with that of similar schools. The report cards had no effect on the immunization rates for the schools that were sent the report cards, compared with the schools that were not sent report cards.^{19,20}

Evaluation 6

Postcards promoting influenza vaccination were mailed to 43,215 patients in the St. Cloud Veterans Affairs Health Care System in Minnesota in a study conducted in 2017–2018. Three different postcards were designed using evidence from behavioral science: a basic postcard providing information about how to get a flu shot, a peer-group-influence postcard noting how many St. Cloud veterans get the shot, and an implementation postcard that prompted veterans to write a concrete plan for getting a shot at a specific time and place. There were no statistically significant differences in the uptake or timing of flu shots among the groups receiving the three postcards.²¹

Evaluation 7

The New York Harbor Veterans Affairs Health Care System sent emails reminding patients to get their flu shots in a study conducted in 2017-2018. A total of 27,162 patients were assigned to either a treatment group or a control group. Using evidence from behavioral studies, messages sent to the treatment group framed getting a flu shot as a default course of action (requiring the patient to take action to opt out); gave an implementation prompt; and presented the benefits of a shot as being concrete and realized in the near term, providing protection within two weeks. The control group did not receive any emails. The emails had no significant effect on the uptake or timing of flu shots.22

Evaluation 8

After a redesign, the Atlanta Veterans Affairs Health Care System's EHR system bundled together three vaccination reminders to clinicians, provided an immunization information dashboard for each patient, and shared talking points that providers could use to address patient refusal or vaccine hesitancy. The evaluation, conducted in 2018–2019, enrolled 84 primary care team clusters that saw 28,941 unique patients during the test period. The difference in vaccination rates between the patients seen by providers exposed to the redesign and those seen by providers not exposed to the redesign was statistically insignificant.^{23,24}

In summary, two of the eight individual evaluations yielded statistically significant effects. In Evaluation 1, letter reminders about influenza vaccination sent to older Medicare beneficiaries increased the probability that they would get an influenza vaccination by 0.4 to 0.7 percentage points (depending on the version of the letter)-a mean of 0.59 percentage pointsrelative to a group who received no reminder letter.¹³ In Evaluation 4, postcard reminders sent to Louisiana residents ages 65-70 years in October increased the number of influenza, tetanus, pneumococcal, and shingles vaccinations they received (analyzed together) by a statistically significant 0.27 percentage points. However, later postcards mailed to different groups did not have a detectable effect. The overall difference in vaccination rates between postcard and no-postcard groups was smaller than 0.27 but still statistically significant: 0.16 percentage points.¹⁷

To gain insights for COVID-19 vaccination campaigns from the OES studies, we performed a *meta-analysis*—a statistical analysis aggregating data from a group of related studies—of the six evaluations that measured vaccination rates at the individual level. We conducted the meta-analysis using a single number representing the effect size from each of those six evaluations (that is, the 0.59 and 0.16 percentage points corresponding to the average treatment condition effects in Evaluations 1 and 4). See the Supplemental Material for technical details.

The meta-analysis indicated that the effect from the communications was small and not statistically significant. We based this conclusion on the confidence interval we calculated. A confidence interval is determined using a procedure that gives a range of values that contains the true effect size some proportion of the time. For instance, if this meta-analysis were repeated 100 times with different data, 95 of those times the 95% confidence interval that we calculated would contain the true size of the effect in the sampled population. For the OES vaccination uptake evaluations, the 95% confidence interval

Evaluation	Collaborator	Sample size	Vaccine type	Population	Year(s)	Outcome	Treatment condition
1	Centers for Medicare and Medicaid Services	227,955	Influenza	Medicare beneficiaries 66+ years	2014–2015	Vaccination uptake	One of four letters encouraging flu vaccination
2	Duke University Health System	2,002	Influenza	Pregnant women	2016-2017	Vaccination uptake	Targeted EHR message on the flu vaccine
3	National Vaccine Program Office	591,221	Influenza, whooping cough	Potentially pregnant women	2017	Ad click rates	One of four variations of ads highlighting maternal immunization
4	Louisiana Department of Health	208,867	Numerous	Adults (65–70 years) overdue for at least one of four vaccines	2017-2018	Vaccination uptake	A postcard reminder sent in October, November, or December
5	City Department of Health	700 schools	All required childhood vaccines	School and daycare center leadership	2017-2018	Vaccine compliance	A vaccine compliance report card
6	St Cloud Veterans Affairs	43,215	Influenza	Veterans 18+ years	2017-2018	Vaccination uptake	One of two postcards informed by insights from the behavioral sciences
7	New York Harbor Veterans Affairs	27,162	Influenza	Veterans 18+ years	2017–2018	Vaccination uptake	Email encouraging flu vaccination and providing action- relevant information
8	Atlanta Veterans Affairs	28,941	Influenza, pneumococcal, Tdap	Veterans 18+ years	2018–2019	Vaccination uptake, all appointments in study period	Primary care teams received modified clinical reminders in the EHR system, vaccination dashboard, and suggested talking points

Table 1.	Key proj	ect characteristi	cs of Office of	f Evaluation Sci	iences va	ccination uptake	evaluations,
includin	g primary	collaborators,	project conte>	kt, evaluation d	lesign, & l	key findings	

Note. EHR = electronic health record; N/A = not applicable. Cost estimates refer to the *ongoing marginal cost*—the cost of delivering an intervention to a target population as an addition to a preexisting program—based on assumptions about the relative costs of these various distribution types. In many cases, the communications could be sent using existing systems, so the marginal cost was zero or very low. This cost framework is discussed in more detail in reference 32. For evaluations that looked at vaccination compliance or updates, means for treatment and control groups are average vaccination rates calculated based on raw data. An exception is Evaluation 4, which had gaps in data availability, and the mean had to be estimated using a statistical model. Means for Evaluations 3 and 5, meanwhile, capture the average click rate on ads and likelihood of immunization compliance, respectively. For further information about Evaluation 1, see references 13 and 14; Evaluation 2, see reference 15; Evaluation 3, see reference 16; Evaluation 4, see references 17 and 18; Evaluation 5, see references 19 and 20; Evaluation 6, see reference 21; Evaluation 7, see reference 22; and Evaluation 8, see reference 23.

Evaluation	Control condition	Treatment mean	Control mean	Key findings	Cost
1	No letter	26.5% (across treatments)	25.9%	Any letter compared with no letter statistically significantly increased vaccination rates by 0.59 percentage points. A letter from the Surgeon General generated the largest effects.	Low
2	No targeted EHR message	38.3%	40.1%	The targeted message generated a statistically insignificant (1.5-percentage-point) drop in flu vaccine uptake.	Very low
3	Ad variations (no one comparison group)	0.15%–0.16% (across treatments)	N/A	The ads had no differential impact on click-through rates.	No cost
4	January postcard	8.75% (across treatments)	8.59%	The October reminder had a small but statistically significant effect (0.27 percentage points), whereas postcards sent later had no effect.	No cost
5	No report card	76.3%	76.2%	The report card did not increase immunization compliance at treated schools compared with control schools.	Moderate
6	Basic (not behaviorally informed) postcard	40.0%	40.1%	The postcards informed by insights from the behavioral sciences generated a combined statistically insignificant (0.4 percentage point) drop in vaccine uptake.	No cost
7	No email	20.3%	20.2%	The email message generated a statistically insignificant increase (0.4 percentage points) in vaccination uptake and also did not affect vaccination timing.	Very low
8	Status quo EHR system	20.74%	19.18%	The EHR intervention generated a statistically insignificant increase (1.6 percentage points) in vaccination rates among treated patients.	Multiple/ unknown

Table 1. Key project characteristics of Office of Evaluation Sciences vaccination uptake evaluations, including primary collaborators, project context, evaluation design, & key findings (*continued*)

_

for the difference in the vaccination rate in the treatment versus control conditions ranged from -0.004 (virtually no effect) to 0.394 percentage points. In other words, interventions like those in the OES evaluations are likely to reliably generate effects of no more than about half a percentage point.

Lessons for the COVID-19 Era

We draw four main lessons from our review of the OES evaluations.

Lesson 1

The first lesson is that behaviorally informed direct communications can increase vaccination rates at scale but may have smaller, less reliable effects than much of the published literature suggests.

The OES evaluations provide ballpark estimates for the effects that behaviorally informed direct communications might have at scale. Although the mailed reminders that yielded statistically significant effects in two studies produced small increases in the percentages of people who got vaccinated, those small differences translated into thousands of additional vaccinations, which may be considered meaningful by program managers.

Still, a public health official planning a vaccination campaign to combat COVID-19 or another disease would want to be mindful of the small sizes of the effects. A review of published studies gives the impression that direct messaging to individuals is more effective than the OES's largescale, real-world evaluations indicate is the case.

We have several reasons for putting more stock in the OES evaluations' findings of small effects than in results from the studies described in the wider literature, including the studies that motivated OES and its collaborators to undertake the scaled-up interventions. For one thing, in contrast to the six OES evaluations that measured actual vaccination uptake, much of the literature applying behavioral science to vaccination focuses on individuals' thoughts and feelings about vaccinations rather than actual vaccination uptake.⁶ It is common for published studies to measure the likelihood of vaccination in a hypothetical scenario or an individual's intention to be vaccinated rather than actual vaccination uptake. (See the 2011 study by Punam Anand Keller and her colleagues for an example of using a hypothetical scenario.)²⁵ But people often fail to follow through on their intentions to act.²⁶

Two non-OES studies that randomly assigned communications and measured actual vaccination rates, albeit with sample sizes under 10,000 participants, found effects in the 2- to 4-percentage-point range.^{11,12} A systematic review of studies exploring the efficacy of emailed reminders to vaccinate found increases in vaccine uptake ranging from 2 to 11 percentage points for people sent an email compared with people who were sent no reminder.²⁷

Another reason to trust the OES evaluation findings is that, as we noted earlier, the median sample size of 55,000 across the eight OES evaluations is considerably larger than that reported in most published studies. Finally, OES reported on the results of every evaluation it conducted.

A closer look at the OES results suggests that the context in which communications are used may explain why some effects seen in studies are not easy to replicate in government evaluations. For example, the OES evaluations did not find the effect seen in one recent field experiment done in an urban health clinic system. In that experiment, researchers testing 19 different text-message vaccination reminders in a sample of roughly 47,000 patients found an average increase of 2.1 percentage points in vaccination uptake.²⁸ The text messages were sent by primary care providers to a sample of patients who had upcoming appointments. The difference between that context and the government sending letters to older adults on Medicare (for instance) may provide a partial explanation for the smaller effects observed in the OES evaluations. The health clinic sample, consisting of people who had appointments scheduled with a familiar health care provider, may have been more responsive to messaging than were the OES evaluation's sample of older adults on Medicare. In addition, working at large scale and in a government context sometimes affects which elements of a messaging campaign can be included. We discuss this point in more detail in Lesson 3.

The finding that behaviorally informed direct communications are likely to have only small effects at scale highlights the importance of sample size in a randomized controlled trial to evaluate the efficacy of such interventions. In many cases, a randomized control trial needs quite a large sample (several thousands of people) to achieve sufficient power to detect effects in a real-world context that has many additional influences on behavior that can reduce the salience and effect of an intervention.

Lesson 2

The second lesson is that additional evidence is needed to evaluate how the cost-effectiveness of behaviorally informed direct communications compares with the cost-effectiveness of other interventions.

Arguments in favor of using communication strategies to influence behavior tend to emphasize that these are inexpensive to implement when calculating costs on a per-recipient basis. Light-touch approaches like direct communications are generally seen as having a low cost per participant and being easy to implement relative to heavier-handed approaches like redesigning forms, prescheduling appointments, or offering material incentives. Also, direct communications can be aimed more precisely at particular individuals or subgroups than is possible with some alternatives, such as commercial advertising campaigns.

Only a few researchers have compared the cost-effectiveness of behaviorally informed communication interventions with the cost-effectiveness of approaches such as financial incentives or policy mandates.²⁹⁻³¹ These studies generally find that communications compare favorably to other approaches. Similarly, a published report on one of the OES vaccination uptake evaluations¹³ extrapolated from the cost of printing and sending letters to argue that

the cost per additional vaccination in the most effective treatment condition was approximately \$90, in line with costs of other approaches. The small effect sizes in the OES evaluations highlight the importance of determining whether the costs of various approaches are justified by the likely outcomes.

To date, OES vaccination uptake evaluations have not collected comprehensive cost information, including hours and salary costs for those involved in delivering an intervention. However, OES recently developed a framework to roughly categorize interventions based on their approximate ongoing marginal costthe added cost of delivering the intervention along with other communications.³² Using this framework, the eight vaccination uptake interventions evaluated by OES include three with no cost (defined as involving no new change to a delivery medium already in use), two at very low cost (from added e-mail), one at low cost (from added printing, printing and mailing, or phone messages), one at moderate cost (from added staffing costs as part of intervention delivery), and one with costs labeled "multiple or unknown" (from the use of more than one of the changes listed above or from other interventions, such as redesigned EHR messaging). The small effect sizes observed for behaviorally informed direct communications suggest that it may be most sensible to deploy these interventions when they can be delivered at very low or no cost, such as by using an existing communication pipeline.

To build stronger evidence about cost-effectiveness, future research needs to record more comprehensive cost data.³³ Ideally, researchers would go beyond printing and mailing costs, capturing both the administrative costs to design and deliver such interventions and the burdens the interventions impose on recipients.^{34,35} For example, one possible comparison is between behaviorally informed direct communications and material incentives.³⁶ Several studies have found that monetary payments increased vaccination rates,^{37,38} although, as Tom Chang and his coauthors have reported, that is not always the case.³⁹ If payments have orders-of-magnitudelarger effects on vaccination, they may actually be more cost-effective than are direct communications that cost less per target. Additionally, if these strategies have different effects on the behavior of nonidentical groups of people, it may be cost-effective to use both approaches in parallel.

Lesson 3

The third takeaway is that rapid evaluations of vaccination uptake interventions in real-world contexts are essential for learning what works in specific contexts for populations of interest.

The OES vaccination portfolio testifies to the importance of evaluating interventions as they are deployed in the field. As might be expected, both implementation details and effect sizes appear to depend highly on context, so engaging in testing during study implementation and designing studies to have high statistical power are both essential. If vaccination campaigns are staged to incorporate rapid evaluations of different approaches rather than deployed as a systemwide rollout of a single strategy, investigators will be able to quickly (and relatively cheaply) discard approaches that are not working and tweak their efforts based on observed results,^{40,41} enabling vaccination efforts to become increasingly effective over time. Widespread and rapid randomized controlled trials of vaccination uptake interventions could enable the COVID-19 vaccination campaign to build evidence about how much (if at all) interventions work to increase vaccination rates.

An important contribution of the OES vaccination portfolio is its demonstration that when scaling up the best practices outlined in the research literature, investigators may find that practical constraints dilute the expected effects of an intervention. For example, OES drew on a study in which about 3,200 utility company employees were sent a letter listing the days, times, and location of a workplace vaccination clinic.¹² The letters sent to employees in the treatment groups included a planning prompt that encouraged them to write in the date or date and time on which they planned to get their shot. OES added similar planning prompts to some of the letters sent to approximately 228,000 Medicare beneficiaries in Evaluation 1, which was described earlier in this article,¹³ but it was not feasible to include information about the locations and hours of operation of local vaccination clinics. The OES study produced a smaller increase in vaccination uptake, which suggests that including a clinic's location and hours might be necessary to reap the full benefit of a planning prompt. Issues of this sort may only become evident when a strategy is evaluated in the context in which it will be applied.

A second example of the practical constraints that can be revealed by real-world tests comes from Evaluation 8, which issued reminders to clinicians in Atlanta through a revamped EHR system at the Atlanta Veterans Affairs Health Care System, bundled patients' needed vaccinations together, and provided talking points for clinicians to use to encourage vaccination. In an earlier study, Amanda F. Dempsey and her colleagues had tested an intervention that included providing 2.5 hours of training to providers in how to use language that presumes patients have a plan to receive the human papillomavirus vaccination rather than initiating a discussion about options.⁴² That study found a 9.5-percentage-point increase in the initiation of a human papillomavirus vaccine series (see also a study that involved a one-hour training session⁴³). Building on that approach, OES modified an EHR system to encourage providers to use language that presumed the patient would vaccinate (for example, "It is time for your X shot today"). The change was part of a suite of modifications to the EHR system designed to make it easier for providers to recommend and order vaccines. Subsequent conversations with providers in the OES evaluation indicated that many did not actually use the presumptive language that was suggested.24 This implementation information is invaluable for informing the design of future interventions, which might try alternative communication strategies or use intensive training.

Lesson 4

The final takeaway is that leveraging existing vaccination administration systems to support randomized evaluations can make evidence

building easier and enable practitioners to tweak vaccine programs for maximum effectiveness.

The OES vaccination portfolio demonstrates the value of working within vaccination administration systems that can support randomized evaluations.44 These studies were conducted guickly (often within a single influenza season) and at low cost by making behaviorally informed design changes to the content or delivery schedule of existing communication programs, which then delivered variants to randomly selected recipients through the existing systems. OES projects show that randomized evaluation can be embedded in a variety of systems with differing data capabilities, even within complex administrative systems ranging from a city department of health to a regional Veterans Affairs health care system. A system need not be specially designed for randomized controlled trials to enable randomized evaluations. It would be particularly easy to evaluate vaccination strategies on a national scale if there were a single federal immunization registry that recorded the vaccination status of every individual or if existing local immunization registries were standardized, which would enable the identification and random assignment of potential vaccination recipients to interventions.

The OES evaluations measured outcomes at low cost by using existing administrative data, such as that captured by state immunization registries, EHRs, and medical claims databases. The more comprehensive and up-to-date the databases, the more useful they are for measuring outcomes in evaluations. For instance, the availability of real-time data about pediatric vaccinations was crucial for the success of the OES collaboration with the city health department in Evaluation 5 because it facilitated the introduction of up-to-date immunization compliance report cards for schools. In contrast, the Louisiana Department of Health postcard collaboration, Evaluation 4, was complicated by the fact that health care providers are not required to report adult vaccinations.

Conclusion

The success of efforts to combat COVID-19 will depend critically on whether people get vaccinated. Communications are a key tool that governments can use to encourage vaccination. Together, eight randomized evaluations of efforts to increase routine vaccinations show that direct communications may increase vaccination uptake, but effect sizes are small. The small effects imply that such communications are a complement to but not a substitute for vaccination policies and programs that maximize convenience and access—for example, the widespread availability of free vaccinations, perhaps with incentives or mandates.

It is worth considering how the context of COVID-19 vaccinations may differ from the context for influenza and other routine vaccinations. Communications that increase the uptake of influenza and other common vaccines typically do so by reminding people who may otherwise forget to get vaccinated to do so and making it easier for them to follow through on existing intentions.⁶ One review described this as "leveraging, but not trying to change, what people think and feel."⁶ These interventions are typically deployed in situations where vaccine supply exceeds demand.

Reported increases in vaccine hesitancy and resistance in recent years likely will create new challenges. Regardless of the specific challenges for continuing COVID-19 vaccination campaigns, the initial demand for COVID-19 vaccinations in the United States exceeded supply. By the beginning of 2022, the situation had reversed in the United States, and hesitancy and resistance to vaccination were reported at home and abroad.45 The OES evaluations show how vaccination uptake interventions can be rapidly and rigorously evaluated at a large scale. Planning for these evaluations now and deploying them soon will allow for the collection of much-needed evidence about how to best apply communications and other interventions as part of current and future vaccination efforts.

author affiliation

Kappes: U.S. General Services Administration's Office of Evaluation Sciences and London School of Economics and Political Science. Toma: U.S. General Services Administration's Office of Evaluation Sciences and Harvard University. Balu: U.S. General Services Administration's Office of Evaluation Sciences and MDRC. Burnett: U.S. General Services Administration's Office of Evaluation Sciences. Chen: U.S. General Services Administration's Office of Evaluation Sciences and University of Illinois at Urbana-Champaign. Johnson: U.S. General Services Administration's Office of Evaluation Sciences and Dartmouth College. Leight: U.S. General Services Administration's Office of Evaluation Sciences and International Food Policy Research Institute. Omer: U.S. General Services Administration's Office of Evaluation Sciences and Yale Institute for Global Health. Safran: U.S. General Services Administration's Office of Evaluation Sciences. Steffel: U.S. General Services Administration's Office of Evaluation Sciences and Northeastern University. Trump: U.S. General Services Administration's Office of Evaluation Sciences and University of Memphis. Yokum: U.S. General Services Administration's

Office of Evaluation Sciences and Brown University. Debroy: U.S. General Services Administration's Office of Evaluation Sciences. Corresponding author's email: pompa.debroy@ gsa.gov.

author note

This work was funded in part by the National Vaccine Program Office and the Laura and John Arnold Foundation. We thank the many federal agencies, collaborators, academic affiliates, and the Office of Evaluation Sciences (OES) team members involved in developing and implementing the OES vaccination portfolio and the eight randomized evaluations referenced. A list of some of these collaborators is in the Supplemental Material at http://behavioralpolicy.org/ journal. We granted unlimited and unrestricted rights to the General Services Administration to use and reproduce all materials in connection with the authorship.

supplemental material

- http://behavioralpolicy.org/journal
- Acknowledgments
- Additional Figure

references

- 1. U.S. Department of Health and Human Services. (2022, March 22). *COVID-19 vaccines*. https://www.hhs. gov/coronavirus/covid-19-vaccines/ distribution/index.html
- National Vaccine Advisory Committee. (2020). 2020 national vaccine plan development: Recommendations from the National Vaccine Advisory Committee. *Public Health Reports*, 135(2), 181–188. https://doi. org/10.1177/0033354920904074
- Norris, T., Vahratian, A., & Cohen, R. A. (2017). Vaccination coverage among adults aged 65 and over: United States, 2015 (NCHS Data Brief 281). U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. https://www.cdc.gov/ nchs/data/databriefs/db281.pdf
- Tacken, M. A. J. B., Jansen, B., Mulder, J., Campbell, S. M., & Braspenning, J. C. C. (2015). Dutch influenza vaccination rate drops for fifth consecutive year. *Vaccine*, 33(38), 4886–4891. https://doi. org/10.1016/j.vaccine.2015.07.052
- 5. lacobucci, G. (2019). Child vaccination rates in England fall across the board, figures show. *The BMJ, 366*, Article I5773. https://doi.org/10.1136/bmj.I5773
- Brewer, N. T., Chapman, G. B., Rothman, A. J., Leask, J., & Kempe, A. (2017). Increasing vaccination: Putting psychological science into action. *Psychological Science in the Public Interest*, 18(3), 149–207. https://doi. org/10.1177/1529100618760521
- Chen, F., & Stevens, R. (2017). Applying lessons from behavioral economics to increase flu vaccination rates. *Health Promotion International*, 32(6), 1067– 1073. https://doi.org/10.1093/heapro/ daw031
- Chou, W.-Y. S., Burgdorf, C. E., Gaysynsky, A., & Hunter, C. M. (2020). COVID-19 vaccination communication: Applying behavioral and social science to address vaccine hesitancy and foster vaccine confidence. National Institutes of Health. https://obssr.od.nih. gov/wp-content/uploads/2020/12/ COVIDReport_Final.pdf
- 9. DellaVigna, S., & Linos, E. (2022). RCTs to scale: Comprehensive evidence from two nudge units. *Econometrica*, 90(1), 81–116. https://doi.org/10.3982/ ECTA18709
- Franco, A., Malhotra, N., & Simonovits, G. (2014, August 28). Publication bias in the social sciences: Unlocking the file

drawer. Science, 345(6203), 1502–1505. https://doi.org/10.1126/science.1255484

- Stockwell, M. S., Kharbanda, E. O., Martinez, R. A., Vargas, C. Y., Vawdrey, D. K., & Camargo, S. (2012). Effect of a text messaging intervention on influenza vaccination in an urban, low-income pediatric and adolescent population: A randomized controlled trial. JAMA, 307(16), 1702–1708. https://doi. org/10.1001/jama.2012.502
- Milkman, 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(26), 10415–10420. https://doi. org/10.1073/pnas.1103170108
- 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*(10), 743–749. https://doi. org/10.1038/s41562-018-0432-2
- 14. Office of Evaluation Sciences. (2018). Encouraging flu vaccine uptake among Medicare beneficiaries. U.S. General Services Administration. https://oes.gsa. gov/assets/abstracts/15xx-medicareflu-vaccine-uptake-abstract.pdf
- 15. Office of Evaluation Sciences. (2018). Increasing flu shot uptake among pregnant women. U.S. General Services Administration. https://oes.gsa.gov/ assets/abstracts/1735-flu-shot-abstract. pdf
- Office of Evaluation Sciences. (2018). Testing variations of maternal immunization messages. U.S. General Services Administration. https://oes.gsa. gov/assets/abstracts/1736-pregnancyvaccines-ads-abstract.pdf
- Chen, N., Trump, K.-S., Hall, S., & Le, Q. (2023). The effect of postcard reminders on vaccinations among the elderly: A block-randomized experiment. *Behavioural Public Policy*, 7(2), 240–265. https://doi.org/10.1017/bpp.2020.34
- 18. Office of Evaluation Sciences. (2018). Increasing vaccine uptake among seniors. U.S. General Services Administration. https://oes.gsa.gov/ assets/abstracts/1738-increasingvaccine-uptake.pdf
- 19. Office of Evaluation Sciences. (2018). Increasing immunization compliance in schools. U.S. General Services Administration. https://oes.gsa.gov/ assets/abstracts/1737-vaccine-reportcards.pdf

- Leight, J., & Safran, E. (2019). Increasing immunization compliance among schools and day care centers: Evidence from a randomized controlled trial. *Journal of Behavioral Public Administration*, 2(2). https://doi. org/10.30636/jbpa.22.55
- 21. Office of Evaluation Sciences. (2018). Increasing flu shot uptake among St. Cloud veterans. U.S. General Services Administration. https://oes.gsa.gov/ assets/abstracts/1740-flu-shots-va-stcloud.pdf
- 22. Office of Evaluation Sciences. (2018). Increasing flu shot uptake among New York Harbor veterans. U.S. General Services Administration. https://oes.gsa. gov/assets/abstracts/1743-flu-shots-vanew-york-harbor.pdf
- 23. Office of Evaluation Sciences. (2019). Increasing vaccine uptake among Atlanta veterans. U.S. General Services Administration. https://oes.gsa.gov/ assets/abstracts/1803-vaccine-uptakeatlanta-va-abstract.pdf
- Debroy, P., Balu, R., Burnett, R., Johnson, R., Kappes, H. B., Wallace, J. M., & Marconi, V. C. (2023). A cluster randomized controlled trial of a modified vaccination clinical reminder for primary care providers. *Health Psychology*, *42*(3), 195–204. https://doi. org/10.1037/hea0001218.
- 25. Keller, P. A., Harlam, B., Loewenstein, G., & Volpp, K. G. (2011). Enhanced active choice: A new method to motivate behavior change. *Journal of Consumer Psychology*, 21(4), 376–383. https://doi. org/10.1016/j.jcps.2011.06.003
- 26. Sheeran, P., & Webb, T. L. (2016). The intention–behavior gap. Social and Personality Psychology Compass, 10(9), 503–518. https://doi.org/10.1111/ spc3.12265
- Frascella, B., Oradini-Alacreu, A., Balzarini, F., Signorelli, C., Lopalco, P. L., & Odone, A. (2020). Effectiveness of email-based reminders to increase vaccine uptake: A systematic review. *Vaccine*, *38*(3), 433–443. https://doi. org/10.1016/j.vaccine.2019.10.089
- Milkman, K. L., Patel, M. S., Gandhi, L., Graci, H. N., Gromet, D. M., Ho, H., Kay, J. S., Lee, T. W., Akinola, M., Beshears, J., Bogard, J. E., Buttenheim, A., Chabris, C. F., Chapman, G. B., Choi, J. J., Dai, H., Fox, C. R., Goren, A., Hilchey, M. D., . . . Duckworth, A. L. (2021). A megastudy of text-based nudges encouraging patients to get vaccinated at an upcoming doctor's appointment. *Proceedings of the National Academy of Sciences, USA*, *118*(20), Article e2101165118. https://doi. org/10.1073/pnas.2101165118

- Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., Tucker-Ray, W., Congdon, W. J., & Galing, S. (2017). Should governments invest more in nudging? *Psychological Science*, 28(8), 1041–1055. https://doi. org/10.1177/0956797617702501
- Patterson, R. W., & Skimmyhorn, W. L. (2021). How do behavioral approaches to increase savings compare? Evidence from multiple interventions in the U.S. Army. TIAA Institute. https:// www.tiaa.org/content/dam/tiaa/ institute/pdf/insights-report/2021-03/ tiaa-institute-how-do-behavioralapproaches-to-increase-savingscompare-ti-skimmyhorn-march-2021. pdf
- Wagner, Z., Montoy, J. C. C., Drabo, E. F., & Dow, W. H. (2020). Incentives versus defaults: Cost-effectiveness of behavioral approaches for HIV screening. *AIDS and Behavior*, 24(2), 379–386. https://doi.org/10.1007/ s10461-019-02425-8
- 32. Office of Evaluation Sciences. (2021). What does it cost to incorporate behaviorally informed interventions within government programs? U.S. General Services Administration. https:// oes.gsa.gov/blog/cost-analysis/
- Dhaliwal, I., Duflo, E., Glennerster, R., & Tulloch, C. (2013). Comparative cost-effectiveness analysis to inform policy in developing countries: A general framework with applications for education. In P. Glewwe (Ed.), *Education policy in developing countries* (pp. 285–338). University of Chicago Press. https://doi.org/10.7208/ chicago/9780226078854.003.0008
- Damgaard, M. T., & Gravert, C. (2018). The hidden costs of nudging: Experimental evidence from reminders in fundraising. *Journal of Public Economics*, 157, 15–26. https://doi. org/10.1016/j.jpubeco.2017.11.005
- 35. Goldstein, D. G., Suri, S., McAfee, R. P., Ekstrand-Abueg, M., & Diaz, F. (2014). The economic and cognitive costs of annoying display advertisements. *Journal of Marketing Research*, 51(6), 742–752. https://doi.org/10.1509/ jmr.13.0439
- 36. Banerjee, A., Chandrasekhar, A. G., Dalpath, S., Duflo, E., Floretta, J., Jackson, M. O., Kannan, H., Loza, F. N., Sankar, A., Schrimpf, A., & Shrestha, M. (2021). Selecting the most effective nudge: Evidence from a large-scale experiment on immunization (NBER Working Paper 28726). National Bureau of Economic Research. https://doi. org/10.3386/w28726

- Bronchetti, E. T., Huffman, D. B., θ Magenheim, E. (2015). Attention, intentions, and follow-through in preventive health behavior: Field experimental evidence on flu vaccination. Journal of Economic Behavior & Organization, 116, 270–291. https://doi.org/10.1016/j. jebo.2015.04.003
- Mantzari, E., Vogt, F., & Marteau, T. M. (2015). Financial incentives for increasing uptake of HPV vaccinations: A randomized controlled trial. *Health Psychology*, *34*(2), 160–171. https://doi. org/10.1037/hea0000088
- 39. Chang, T., Jacobson, M., Shah, M., Pramanik, R., & Shah, S. B. (2021). Financial incentives and other nudges do not increase COVID-19 vaccinations among the vaccine hesitant (NBER Working Paper 29403). National Bureau of Economic Research. https://doi. org/10.3386/w29403
- 40. Hadad, V., Rosenzweig, L. R., Athey, S., & Karlan, D. (2021). *Practitioner's guide: Designing adaptive experiments.* Stanford Graduate School of Business Golub Capital Social Impact Lab; Northwestern Kellogg School of Management. https://www.gsb.stanford. edu/faculty-research/publications/ practitioners-guide-designingadaptive-experiments
- Kasy, M., & Sautmann, A. (2021). Adaptive treatment assignment in experiments for policy choice. *Econometrica*, 89(1), 113–132. https:// doi.org/10.3982/ECTA17527
- 42. Dempsey, A. F., Pyrznawoski, J., Lockhart, S., Barnard, J., Campagna, E. J., Garrett, K., Fisher, A., Dickinson, L. M., & O'Leary, S. T. (2018). Effect of a health care professional communication training intervention on adolescent human papillomavirus vaccination: A cluster randomized clinical trial. *JAMA Pediatrics*, *172*(5), Article e180016. https://doi.org/10.1001/ jamapediatrics.2018.0016
- 43. Brewer, N. T., Hall, M. E., Malo, T. L., Gilkey, M. B., Quinn, B., & Lathren, C. (2017). Announcements versus conversations to improve HPV vaccination coverage: A randomized trial. *Pediatrics*, 139(1), Article e20161764. https://doi.org/10.1542/ peds.2016-1764
- 44. Finkelstein, A. (2020). A strategy for improving US health care delivery— Conducting more randomized, controlled trials. New England Journal of Medicine, 382(16), 1485–1488. https://doi.org/10.1056/NEJMp1915762

45. Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour*, 5(3), 337–348. https://doi. org/10.1038/s41562-021-01056-1

Drawing what I value





Can a visual valuesaffirmation intervention improve test scores of students in areas affected by crisis?

Daniel D. Shephard, Ali Osseiran, & Fadi Makki

abstract

Values-affirmation (VA) exercises, which direct people's attention to aspects of their lives that they value and broaden their sense of self, have been shown to improve performance in many populations, particularly those who worry that doing poorly will feed into negative stereotypes of the ethnic or other social groups they belong to. Most studies of VA have examined its benefits in highly literate, economically stable, Englishspeaking populations and have used written exercises. We conducted a randomized controlled trial of a visual VA exercise in an understudied population: marginalized Arabic-speaking students (mostly Syrians) living in a context (Lebanon) affected by conflict. Before taking final exams for a program to improve basic Arabic and English literacy skills and math proficiency, the participants, ages 14-24 years, made a drawing that represented a value important to them. This visual VA exercise improved performance on the Arabic test, particularly among the Syrians, suggesting that, at least for the Arabic test, it reduced anxiety related to stereotyping, allowing students to relax enough to demonstrate their true ability. If replicated, our findings would suggest that schools could use such exercises to improve the value of test scores for guiding decisions about next steps in the education of marginalized students in a context affected by conflict.

Shephard, D. D., Osseiran, A., & Makki, F. (2023). Can a visual values-affirmation intervention improve test scores of students in areas affected by crisis? *Behavioral Science & Policy*, 9(1), 27–42. https://doi.org/10.XXXX/xxxxx

esting is increasingly used to rank students, schools, and countries' educational systems,¹ and students know that poor results can have distressing repercussions, such as the humiliation of having to repeat a grade.² The pressure to succeed is particularly stressful for students who are marginalized-that is, deprived of social acceptance or resources available to others-as is often the case for ethnic minorities and migrants. Marginalized students often fear that their poor performance will confirm negative stereotypes about their group, such as that group members are less intelligent than nonmarginalized students are.³ This anxiety, in turn, can impair performance and amplify existing achievement gaps between marginalized and nonmarginalized students,^{4,5} an effect that feeds concern that educational systems are replicating inequalities rather than addressing them.6-10

Brief social-psychological interventions that target students' thoughts, feelings, and beliefs offer a promising way to reduce achievement gaps in school. *Values-affirmation* (VA) interventions, in particular, have accumulated a substantial evidence base¹¹⁻¹⁴ and have been applied increasingly in recent years to improve students' performance and reduce gaps in performance between subsets of students.¹⁵⁻¹⁹ These interventions prompt students to bring to mind a value that is important to them, such as having positive relationships with family members or friends or a commitment to religion.

So far, VA interventions have been studied predominantly in highly literate populations, and little is known about their effectiveness in other groups, such as marginalized students living in settings affected by conflict. Yet interventions for these students are sorely needed. They often have poorer educational outcomes than marginalized students in more stable communities do.^{20–22} Furthermore, some have expressed concern that individuals in such populations may be more susceptible to embracing extremism,²³ although the evidence for this is mixed.

We have conducted a randomized controlled trial in Lebanon to directly assess whether a VA intervention could improve the test scores of marginalized students living in a setting affected by the neighboring Syrian conflict and internal crises. Our study addressed this question by creating an Arabic version of a VA exercise and assessing its effect on a population consisting mostly of adolescent and adult refugees from Syria who enrolled in a program in Lebanon that teaches basic Arabic and English literacy and numeracy. Our VA intervention involved having students express their values through drawing rather than in writing, because the literacy level of the students in the program was low. We found that this brief intervention could improve test scores in this population and was more effective for the Syrian students than for others in the program.

Past Research Supporting VA

The classic example of a VA intervention that had a powerful effect on academic performance was tested by Geoffrey L. Cohen and his colleagues, who delivered it to seventh-grade students in the United States attending schools that were in middle- to lower-middle-class areas and that had roughly equal numbers of students of European and African descent.^{11,12} In two randomized controlled trials, participating students were assigned to either treatment or control groups early in the school year. Students assigned to treatment groups were given a list of values and asked to select their most important personal value (in the first study) or their two or three most important values (in the second study); the participants then wrote a paragraph about why the value or values were important to them. Students in the control group for each study were also given the list of values, but they were asked to select their least important value or values from the list, and they wrote a paragraph about why the value or values might be important to someone else.

At the end of the school year, the researchers found an improvement in the grades of the African American students in the treatment groups.¹¹ In the first study, the intervention increased the mean grade point average (GPA) of these pupils by 0.26 on the four-point scale; in the second study, it increased the GPA by 0.34 points. Both results were statistically significant. A follow-up study including students from those two studies and a similar third study conducted with a later group of students found that the effect lasted through the end of the following year, with a gain of 0.33 GPA points.¹² Other studies have demonstrated that VA interventions can increase the probability of college enrollment among African American students seven to nine years later.¹⁴

VA interventions have been shown to improve academic outcomes for other stereotyped social groups, too. For instance, they have improved academic outcomes and increased enrollment in college readiness courses among Latinx middle schoolers,¹⁴ reduced the gender gap between the grades of men and women in a college physics course,¹³ and enhanced performance on tests of mathematical concepts among undergraduate students in a psychology statistics course attended mostly by women.²⁴

Recent studies have replicated the benefits of VA interventions at scale, providing evidence of the validity and robustness of VA interventions in varying settings, albeit with effects that were smaller than those seen in earlier studies. In the United States, a VA intervention delivered to seventh-grade students across an 11-school district led to improved cumulative GPAs among racial and ethnic minority students.²⁵ In the United Kingdom, a study with tenth and eleventh graders across 29 secondary schools found modest effects that were sustained for a year.^{26,27}

Although the many findings supporting the ability of VA interventions to improve academic outcomes are encouraging, the findings are not equally positive in all contexts,²⁸ and several questions remain. For instance, the smaller effect found in the large-scale studies raises the issues of how, when, and for whom VA interventions are most appropriate.^{28–30} In addition, the research to date is marked by three gaps in the populations that have been studied, which may limit the generalizability of the findings to other populations.

First, most studies have been conducted in high-income, English-speaking countries. A search on "values(-) OR self(-) affirmation" in all databases of the Web of Science (https://www. webofscience.com/wos) on April 18, 2021, yielded 1,531 studies, of which 60.9% were in the United States (44.4%), England (10.9%), or Canada (5.6%).

Second, and related to the first point, the populations targeted by most VA studies live in contexts marked by relative stability, even though it is students who are experiencing displacement, military or other conflicts, economic insecurity, or fear for their own or their family's safety who are at heightened risk of low educational attainment and achievement.^{21,22,31,32}

Third, the interventions almost always require study participants to make their affirmations in writing and therefore can only be deployed among populations with high levels of literacy. Indeed, a written intervention with a population that feels it has a low level of literacy may backfire, as it makes such a perceived shortcoming more salient. This risk is not merely theoretical: Studies in which negative stereotypes apart from literacy have been primed (that is, brought to mind) have been shown to negatively affect performance among groups as diverse as African Americans,³³ student athletes,³⁴ older people,³⁵ and immigrants.⁵

Evidence of generalizability is important in light of both the much-publicized concerns over the replicability of many psychological studies³⁶ and the need to better understand the factors that moderate the effectiveness of affirmation interventions.^{28,29}

Despite the predominance of written VA interventions with students, there is reason to suspect that variations that do not rely on writing can be effective. For instance, Crystal C. Hall and her colleagues have found evidence that a verbal self-affirmation exercise used by adults at a soup kitchen improved executive control and fluid intelligence (that is, logical reasoning).³⁷ Moreover, some work has shown that drawing exercises can positively affect emotional states,³⁸⁻⁴⁰ and other research indicates that images can have framing effects (that is, they can define a situation in a way that alters attitudes or behaviors) and priming effects that may be more powerful than those of text

alone.⁴¹⁻⁴³ Finally, a rich literature amassed over more than half a century has examined how individuals' expectations and values affect their visual perceptions⁴⁴ and suggests that some social and cultural groups may be particularly responsive to visual imagery.^{45,46}

Research Into the Mechanisms Underlying How VA Works

Research into the psychological phenomenon of stereotype threat probably helps explain why VA interventions can improve test scores in populations subjected to stereotyping. Stereotype threat occurs when individuals want to perform well on a task, such as a test, but worry that a subpar showing will confirm negative stereotypes of a group that shares their race, ethnicity, nationality, gender, age, or some other aspect of their identity.⁴⁷ This worry, in turn, can increase anxiety and distract the individuals from concentrating fully on the task at hand. For example, women taking a mathematics examination may, in addition to feeling routine anxiety over a test, experience stereotype threat that increases this anxiety,48 because they fear a poor performance will lend credence to the stereotype that women are less skilled than men at mathematics.

The degree to which an individual's performance on a task is impaired because of stereotype threat will be moderated by such factors as whether and how much the task brings the stereotypes to mind (that is, is priming) and how much they feel that their identity is linked to their performance on that type of task. Individuals who care a great deal about performing well in a given domain are more likely to experience the effects of stereotype threat.^{33,49-51} For example, studies have found that when confronted with the stereotype that men perform better than women in math, women who felt that their math proficiency was an important part of their identity were more likely to underperform on math-related tasks then were women who identified less with being adept in that domain.51,52

The effects of priming and of the importance to the individual of doing well on a task were shown in groundbreaking work on stereotype threat by Claude M. Steele and Joshua Aronson,³³ who demonstrated that merely making a racial stereotype salient could impair the performance of Black students relative to White students. Many other studies have found similar results. For instance, scholars have shown the negative effects of stereotype threat on academic performance in racial and ethnic minority groups in general and on other groups in specific domains, such as women in STEM (science, technology, engineering, and mathematics) fields. 33,47,53-56 Hall and her colleagues³⁷ have also shown that stereotype threat related to living in poverty can affect cognitive performance and decision-making in a nonacademic context.

Extrapolation from past research gives reason to believe that refugees and adults students may experience stereotype threat.^{57,58} In the case of refugees, displaced individuals are often viewed in terms of what they lack, which can amplify stereotype threat.⁵⁹ In the case of adults, most of the discourse and materials related to education focus on children. When programs do address adult learners, those programs are framed as being something other than formal education, which reinforces the perceptions among educators and students that adult learners are atypical and have fallen behind others in their age group. Fear of being stereotyped in this way may be particularly strong for adult students who feel that they need to take classes in basic literacy or numeracy.21,22,60

How might VA interventions reduce stereotype threat? Existing research indicates that affirming what one values expands one's perspective on the things that make up one's identity and thus reduces the negative effects of stereotype threat and leads to better outcomes. In other words, broadening individuals' perspective of themselves may enable them to recall that other people's stereotypes do not define them, to trust in their own abilities, and to ease their worry about the consequences of failure.

If this process is a mechanism of change, the effects of a VA would be observed immediately. Accordingly, the timely use of a VA activity
just before academic testing may improve the performance of students who experience stereotyping that they fear will be confirmed by the test. If a VA exercise can ease anxiety related to stereotype threat enough to help reveal a student's true knowledge, it would make testing a more accurate guide to whether stereotyped students need remedial help or are ready to progress and what the next steps in their education should be.

Study Rationale & Predictions

Our study addresses the research gaps described earlier in this article-namely, the lack of research in populations other than English speakers from high-income countries who live in relatively stable conditions and are highly literate. It provides a model for how VA interventions could be deployed in schools among marginalized, low-literacy populations who endure living conditions made unstable by conflict. As we have mentioned, students in these populations and the education systems responsible for educating them are among the populations and contexts most in need of effective, evidence-informed practices to close the achievement and attainment gaps, especially with respect to literacy.^{20,61}

We conducted the study in multiple public academic centers in Lebanon, where most people speak Arabic as their primary language. Lebanon has the world's highest concentration of refugees,⁶² and there are nearly as many refugee students as Lebanese students in the public schools.⁶³ Most Lebanese students attend private schools; the 31% who attend public schools represent the lower socioeconomic segment of society.⁶⁴ The students in this study included adolescents and young adults enrolled in a program intended to remediate gaps in their education.

To the best of our knowledge, this is the first randomized evaluation of a VA intervention in Lebanon specifically and among the Levantine and Gulf countries more generally. It is also the first to include refugees as the majority of participants and thus to contribute to the nascent literature on improving educational outcomes for displaced learners in fragile and conflict-affected contexts. Having fled their homeland, Syrian refugees in Lebanon now find themselves in a country that is itself affected by crises: hosting a large number of refugees (particularly from Syria),65 recovering from conflicts with Israel⁶⁶ and the deadly 2020 explosion of ammonium nitrate in Beirut,⁶⁷ and contending with deepening economic and political problems.⁶⁸ Although all of the students in the academic centers we studied are marginalized to some extent-carrying the stigma of being in the low-income stratum of society and needing remedial education-the Syrians might be considered doubly marginalized in that they are foreigners and do not have the same economic, political, or educational rights as people from Lebanon do.

An additional contribution of this study is that it enabled us to evaluate a visual VA intervention we designed for use in groups whose members' literacy level is low.

The mix of participants in our study allowed us to explore the different effects of our VA intervention in multiple groups whose members commonly experience stereotype threat, including not only refugees but also adult students and females.

Specifically, our study explored the following research questions:

- Can a visual VA intervention lead to an improvement in the test performance of marginalized students with low literacy?
- Will the refugees in our study benefit significantly more from the intervention than the nonrefugees do?
- Will the adult learners in our study benefit significantly more from the intervention than the minors do?
- Will the female learners in our study benefit significantly more from the intervention than the male learners do?

Subject Selection & Study Design

Our sample consisted of individuals aged 14–24 years who were enrolled in an accelerated basic literacy and numeracy program offered at various sites in Lebanon. Those 18 years of age or older are considered adults. The program consisted of two 96-hour cycles (about eight weeks per cycle) covering Arabic literacy, English literacy, mathematics, and information and communication technology skills. The program was open to disadvantaged individuals of all nationalities, but the majority had to be Syrian. Only individuals who had been out of school for at least two years were allowed to enroll.

All students in the program were eligible to participate in the study. Participation was voluntary. Students were given a generic description of the study as a "drawing task" during the consent process. A total of 150 out of the 248 students enrolled in seven learning centers agreed to participate. (See Figure 1 for learning centers' locations and enrollments.) More than half of the participants were female (56.7%); approximately half were adults (46.7%); and almost all of them were Syrian (90.7%), whereas the rest were Lebanese or Palestinian. After they completed the task and their exams, the participants were debriefed and received further details about the intervention and study hypotheses.

Contrary to previous studies, some involving two or more VA interventions throughout the academic year, this study assessed one activity completed on the day of the final exams. As a result, it offers some insight into whether a one-time intervention can be effective.

Participants were randomized on the day of the intervention using a predetermined randomization sequence applied to consenting students who lined up at each center to receive their activity and room assignment. At each center, half of the participants were assigned to the treatment group and half to the control group.



Figure 1. Map of the academic centers $\boldsymbol{\vartheta}$ the number of students enrolled in each center

Note. Of the 248 students enrolled at the centers, 150 participated in the study. The treatment and control groups were evenly matched by gender; percentage of Syrians; percentage of adults; and baseline test scores in math, English, and Arabic. See Table S1 in the Supplemental Material for more details. In both groups, about 90% of the participants were Syrian.

We applied a randomization approach known as a stratified or block strategy to account for systematic differences that might have existed in the populations across centers.

In total, 77 participants were in the treatment group and 73 were in the control group. In both groups, about half the participants were adults and half were female; their baseline Arabic, English, and math scores were equivalent; and about 90% of both groups were Syrian. (See Table S1 in the Supplemental Material for more details.)

Participants in the treatment group were assigned to one room, where they received the VA treatment; those in the control group were assigned to another room and given a placebo activity not expected to affect test performance. The random assignment was completed under the supervision of one or more research assistants to ensure that the sequence was followed and that only individuals with consent forms took part in the study.

After being assigned to their room, participants in the treatment group received two sheets of paper with instructions written in Arabic. The first sheet included a list of 10 values (such as "friendship," "honesty," and "courage"), presented in writing and graphically, along with written instructions telling participants to rank these values. The second sheet contained instructions and a blank space in which participants could illustrate why their top value was important to them. Participants were informed that the quality of the drawing did not matter. A trained research assistant also read the instructions out loud and explained the task to make sure everyone understood it (See Figure 2 for the written instructions and the list of values).

Figure 2. Instructions & list of values presented to participants

Instructions

Treatment group	Control group
List the 10 values from most to least important	List the 10 values from most to least important
to you on a scale of 1 to 10 (1 being the most	to others on a scale of 1 to 10 (1 being the most
important).	important).
Pick the value that is most important to you	Think about your morning routine and try to
(#1) and try to explain why it is important to you	illustrate the first thing you do in the morning
by illustrating it with a drawing. You have 10	with a drawing. You have 10 minutes; you can
minutes; you can draw as much or as little as	draw as much or as little as you want. Don't
you want. Don't worry about the quality of the	worry about the quality of the drawing, it is not
drawing, it is not important!	important!

List of values

English	Arabic
1. Sense of humor	التمتع بحس فكاهى
2. Commitment to religion	الإلتزام بالدين
3. Exercise/sports	ممارسة الرياضية
4. Helping others	مساعدة الآخرين
5. Friendship	الصداقة
6. Family	الرعابة والاهتمام بالعائلة
7. Honesty	الصدق
8. Courage	
9. Creativity	ر م جن رج الارجار بر التفائل
10. Positivity/optimism	الإيجابية ٢ التقاتل

Note. The instructions and list of values were presented in Arabic and are translated into English for readers of this article. The values were also presented graphically.

Participants assigned to the control group ranked the same 10 values according to what they thought others (not what they themselves) valued. They were then asked to draw something from their morning routine. The rest of the procedure matched that used for the treatment group.

The use of a placebo is standard practice in experimental studies of VA.^{28,69} Thus, our use of a placebo control facilitated comparability with previous studies and helped reduce the possibility of our findings being driven by the Hawthorne effect (that is, people modifying their behavior in response to being observed) or simply by the act of drawing. (See note A for a further discussion of placebo use in VA studies.)

The intervention and the placebo activities each lasted 20 minutes. Afterward, all participants were invited to sit for their final exams with the rest of the students in the academic center.

Measures

Placement Test Scores

All students enrolled in the academic centers, whether or not they participated in the study, completed a placement test to assess their baseline Arabic, English, and math skills. The Arabic and math tests were scored on a scale of 0 to 100; the English test was scored on a scale of 0 to 25.

Final Exam Scores

All students, whether or not they participated in the study, completed final exams to assess their performance at the end of the program. The Arabic and math tests were scored on a scale of 0 to 100; the English test was scored on a scale of 0 to 40.

The raw placement and final exam scores of the study participants were retrieved from the centers' administrative records. All test scores were standardized as the percentage of correct answers out of the maximum possible correct answers, to account for the different grading scales used on the placement and exit tests and on the three subject-matter tests. Data regarding the participants' gender, nationality, age, academic center, and grade level were also collected.

Analytic Methodology

The primary analyses were preregistered before the trial was conducted.⁷⁰ The analyses assessed the effect of the VA task on the performance of the participants in the treatment group relative to the performance of the participants in the control group.

Our primary analyses separately examined the effect of the intervention on the test scores for each academic subject and assessed the change in performance relative to the baseline for each participant.

In addition to the primary analyses, we conducted exploratory moderation analyses, which investigated differences in the effect of the intervention on subgroups in our study population—namely, whether the effect of the intervention differed in Syrians versus non-Syrians, adults versus adolescents, and females versus males. (For more details about our statistical methodology, see the Supplemental Material.)

Results

Figure 3 summarizes the results of the primary analyses. Participants in the treatment group scored 6.9 percentage points more on their Arabic test score compared with the participants in the control group, after baseline scores were taken into account (0.069, 95% CI [0.001, 0.137]); this is the equivalent of an effect size of 0.27 standard deviations. (See note B for a discussion of the statistical terms used in this article.) This effect remained when we accounted for the potential influence of age, sex, and nationality (0.070, 95% CI [0.001, 0.139]). However, the intervention had no effect on English or math scores. (See Table S2 in the Supplemental Material for more details.)

The moderation analyses showed that Syrian students benefited more from the intervention than did non-Syrian students with respect to the Arabic test (see Figure 4; for fuller details, also see Table S3 in the Supplemental Material).





Figure 4. Effect of the intervention on the Arabic test scores of Syrians versus non-Syrians



Note. The *y*-axis indicates the effect of the intervention on the proportion of correct answers out of the total number of questions on the Arabic examination. The test scores of the Syrians rose to a statistically significant extent, whereas the drop shown in the non-Syrians' scores was not statistically significant. The error bars represent the 95% confidence intervals using robust standard errors. (See Table S3 in the Supplemental Material for more details.)

"Our study demonstrates the effectiveness of a VA exercise for a new population and context: mostly refugees attending school in a setting affected by crisis."

> When we compared the percentages of correct answers on the Arabic test out of the total possible scores and accounted for variances in baseline scores, we found that the Syrian participants scored 25.6 percentage points higher than the non-Syrian participants did (0.256, 95% CI [0.001, 0.511]). The intervention did not improve Arabic test scores for females versus males or for adults versus youths. We also found no significant influence of nationality or age group on the intervention's lack of effect on math and English test results, nor did we find a significant influence of gender on the math results. We did, however, find that male participants benefited from the intervention more than female participants did when it came to the English test (0.158, 95% CI [0.031, 0.286]). Although only tentative implications can be drawn, this finding may suggest that the intervention would be an effective way to support male students taking English language tests.

Discussion

In a study of primarily Syrian adolescent and adult students enrolled in a basic literacy and numeracy program in Lebanon, we found that a brief, low-cost, visual VA intervention improved study participants' performance on the final exam for the Arabic course. This effect is statistically and substantively important despite null results for study participants' performance on math and English tests. The effect size–0.27 standard deviations—is larger than the aggregate effect size (a Hedge's *g* value of 0.15) reported in a recent meta-analysis examining VA interventions for identity-threatened students.²⁸ Moreover, the effect size is almost twice the average effect size for VA interventions.²⁸ For a sense of scale, consider that in the United States, 0.20 standard deviations is approximately equivalent to the gains in reading between grades 9 and 10. Meanwhile, the difference between student grades in a weak school versus an average school in the United States has been found to generally be between 0.20 and 0.40 standard deviations.⁷¹

The effect size we found may be a conservative estimate of the intervention's effect, because it may have been dampened by several factors. These factors include delivery of the intervention by researchers, the use of the schools' existing exams rather than exams designed specifically to align with previous research, and the immediacy of our follow-up.28 Although the mechanisms behind these dampening effects, which have been found in previous VA studies, are unclear, the following logic could be explored in future research and practice: If students' regular teachers deliver the intervention, students may take the exercise more seriously or find it more meaningful. If the academic outcomes were based on researcher-designed exams, the exams could provide more precise estimates of effects and address more competencies than is possible when existing exams are used. Finally, longer follow-ups in previous studies have been associated with larger effect sizes, even with relatively light-touch VA studies. Conceivably, the difference in outcomes between the treatment and control groups might also have been diminished if the exercise we gave to the control group participants ended up reducing their test anxiety or expanding their conception of self-such as by prompting them to spend a few minutes thinking about something other than the examination and to engage in a relaxing drawing activity.40

Our study expands the literature on VA interventions in three important ways. First, it demonstrates the effectiveness of a VA exercise for a new population and context: mostly refugees attending school in a setting affected by crisis. This is an important contribution, given the lack of robust evidence concerning which interventions can improve the performance of refugee students⁶¹ and the concentration of

previous VA studies geographically in the United States and demographically among nonrefugee racial minorities and women.^{11,12,14,15,72} The lack of gender-based differences in math scores in response to the VA intervention in our study—in contrast to findings in other studies—suggests that such gender-based effects may be influenced by whether gender-related stereotyping is prevalent in the academic contexts in which the intervention is used.⁷³ The moderation-analysis finding that the male students benefited from the intervention more than the female students did with respect to the English exam may reflect the tendency of female learners in Lebanon to outperform male learners in academic subjects.⁷⁴

Second, we show that traditional text-based VA exercises can be replaced by a visual exercise, such as a drawing task, and still be successful. This finding extends the applicability of the technique to low-literacy learners and, potentially, early-grade learners. Being able to use the VA technique with these populations would fill a need, given that low-literacy learners are likely to be stigmatized and are in great need of supportive interventions, especially when they live in areas affected directly or indirectly by conflict.

Third, we provide insight into possible mechanisms of effect for VA interventions. We show that a one-time VA intervention can have an immediate effect on test scores when implemented prior to an examination. This result lends support to the possibility that one of the underlying mechanisms is an expanded sense of self-worth, because an elevation in self-esteem can occur quickly. Moreover, another explanation for improved performance after an intervention—that the student learned new information before an examination—would not be possible, owing to a lack of time for this learning to occur^{28,30,75}

Arguably, VA is likely to broaden students' sense of self, stimulating a feeling of integrity and pride across expanded domains that may otherwise be constricted by stereotype threat. Put differently, engaging in VA likely reminds people that stereotype threat is not all that defines the self, which then minimizes the effect of the threat.

"We show that traditional textbased VA exercises can be replaced by a visual exercise, such as a drawing task, and still be successful."

Future research should examine whether this mechanism is at work by directly analyzing a student's sense of self-worth and stereotype threat before and after a VA intervention. We would also like to see longitudinal research with refugees to determine the intervention's longterm effects in this population.

That our primary analyses revealed an effect on Arabic test scores but not on English or math scores could have several explanations. One has to do with the probability that the students identified more with their results on the Arabic test. Because Arabic is the native language in the region and is linked to cultural and religious values (that is, the students felt high domain identification with the Arabic language), showing oneself to have low proficiency on an Arabic test could be more threatening to an individual's sense of self than would displaying low proficiency in math and/or English-areas that mean less to the students' sense of self (that is, for which they feel low domain identification). Hence, the affirmation of values may have been most effective against anxiety and stereotype threat regarding the Arabic text. Previous research has shown that proficiency in one's native language enhances positive feelings about one's ethnic identity;⁷⁶ the importance of language skills for ethnic identity may be amplified for learners in this region, given the historical importance of the Arabic language in the Mashriq region, which includes Lebanon and Syria.77

Another explanation for the intervention's effect only on Arabic test scores is suggested by differences in the baseline test scores between Syrians and non-Syrians. The Syrian students who benefited from the intervention on their Arabic test had higher baseline levels of performance "We would like to see longitudinal research with refugees to determine the intervention's long-term effects in this population."

> in math and English compared with those of the non-Syrian students. A VA exercise often raises test scores less when students who engage in the exercise have smaller pre-achievement gaps to start with.²⁸

> Last, the self-affirmation intervention may have been more salient and thus more influential during the Arabic portion of the examination because the intervention itself was in Arabic. Future studies should investigate the relationship between the language of the VA intervention and the language of the examination.

> Although our study design and our reliance on administrative data did not allow us to directly test whether a reduction in stereotype threat was a mechanism that enabled better test performance, the greater effect of the intervention on the Arabic test scores of Syrians (as was found by the moderation analyses) provides some suggestive evidence. The finding may indicate that the intervention raised the Syrians' pride in their Arabic identity, an identity linked to the Arabic language, and this rise in pride may have counteracted the anxiety stemming from the stereotyping they experienced as refugees. This mechanism is speculative and would need to be tested in future research; however, our study generates some preliminary evidence that this route of inquiry may be productive.

> What are this study's additional implications for policy and future research? Although our study was small and policy implications should be made cautiously, the results highlight the promise of delivering a brief visual VA intervention to students from low-literacy populations in crisis-affected contexts prior to their taking exams testing the students' knowledge of their

dominant language. As we note in the sidebar Policy & Research Implications, the intervention's greater effect on Syrian participants than on non-Syrian participants also suggests that the intervention could be effective for refugees elsewhere, especially when schooling is disrupted and students need to take a placement test to determine the most appropriate level at which to reenter school.

However, organizations implementing VA interventions should be careful not to risk further stigmatizing a subset of students by publicly limiting the intervention to those individuals. Besides, although VA interventions are designed to be effective with marginalized students, the cumulative evidence does not indicate that they are harmful for the more dominant groups in the same schools.

This study has direct implications only for improving the Arabic test scores of Arabicspeaking students who are in a catch-up program in a crisis context. We encourage organizations working with these and similar populations to extend our findings by combining this visual VA intervention with rigorous research on its effectiveness and investigating how, when, and for whom visual VA interventions are most appropriate. When implementing our intervention in the future, we recommend adjusting it to be delivered by teachers or trainers instead of researchers, as evidence suggests that this approach is more effective.28,78,79 Having the intervention implemented by teachers or trainers would also help to assess its potential effect outside of the experimental context, because those people would likely be the implementers if a visual VA exercise became part of a school's or learning center's standard practices.

As we noted earlier, our study also suggests that Arabic students who engage in a visual VA exercise immediately before taking a test assessing Arabic literacy will enhance their test scores not by learning something new but by reducing their anxiety and stereotype threat or by expanding their sense of self enough to enable them to perform closer to their actual level of proficiency than would otherwise have been the case. After all, they would not have had time to learn something relevant to the test between doing the exercise and starting the test. Longitudinal studies would be needed to determine whether a visual VA intervention can also improve learning trajectories over time, as has been shown in studies of other VA interventions in other populations.^{12,14,26}

We would also like to see replication studies that assess the robustness of the current findings at scale and that are well powered to detect differences in the effects of treatment in various subgroups. Ideally, studies would also investigate how the content of students' drawings relates to the observed effects; recent work shows that the content of essays written in traditional VA exercises can somewhat predict who will benefit most from the intervention.⁸⁰

We hope that our study encourages more researchers and practitioners to explore low-cost social-psychological interventions to support the educational progress of learners with low literacy residing in conflict-affected areas and that it will encourage research into using VA interventions to support low-literacy students more broadly. Such interventions have the potential to benefit millions of learners in fragile countries around the world.

end notes

- A. We are aware of only one VA study (aiming to improve academic performance and course attendance) with both a placebo control group and a business-as-usual group whose participants received no activity to complete apart from what they would normally do at that time. Although experimental imbalances prevented a direct test between the two controls, the evidence suggested that the placebo had, if anything, a positive effect on participants' attendance.⁸¹
- B. Editors' note to nonscientists: Standard deviation is a measure of the amount of variation in a set of values. Approximately two-thirds of the observations fall between one standard deviation below the mean and one standard deviation above the

Policy & Research Implications

- If our findings are replicated, they suggest that visual values-affirmation interventions could help refugees raise their test scores in reading or literacy subjects in their native languages, which are key policy priorities for countries to achieve the United Nations' Sustainable Development Goals for education by 2030.
- Visual values-affirmation interventions conducted prior to academic assessments can enable students to demonstrate their true proficiency in those areas and thereby help educators to gain a better sense of students' abilities and how best to meet their educational needs.
- In populations with low literacy, visual values-affirmation interventions can be effective alternatives to traditional written values-affirmation interventions.
- Future studies to replicate and expand our findings should investigate whether changes in anxiety, stereotype threat, a broadened sense of identity, or some combination of these explains the mechanism by which visual values-affirmation interventions improve performance on academic tests.
- Future studies should also investigate the long-term effects of visual values-affirmation interventions over the course of an entire academic term or year.

mean. Standard error uses standard deviation to determine how precisely one has estimated a true population value from a sample. For instance, if one took enough samples from a population, the sample mean ± 1 standard error would contain the true population mean around two-thirds of the time. Robust standard errors are used instead of typical standard errors when the usual assumptions about the data distribution do not apply. A 95% confidence interval for a given metric indicates that in 95% of random samples from a given population, the measured value will fall within the stated interval.

author affiliation

Shephard: Teachers College, Columbia University. Osseiran: Nudge Lebanon. Makki: Boston Consulting Group and Nudge Lebanon. Corresponding author's e-mail: daniel.shephard@ columbia.edu.

supplemental material

- http://behavioralpolicy.org/journal
- Method & Analysis

references

- 1. Kamens, D. H. (2015). A maturing global testing regime meets the world economy: Test scores and economic growth, 1960–2012. *Comparative Education Review, 59*(3), 420–446. https://doi.org/10.1086/681989
- 2. Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, *30*(3), 420–437. https://doi.org/10.1080/02796015.2001 .12086124
- von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders*, 227, 483–493. https://doi.org/10.1016/j.jad.2017.11.048
- Appel, M., & Kronberger, N. (2012). Stereotypes and the achievement gap: Stereotype threat prior to test taking. *Educational Psychology Review*, 24(4), 609–635. https://doi.org/10.1007/ s10648-012-9200-4
- Appel, M., Weber, S., & Kronberger, N. (2015). The influence of stereotype threat on immigrants: Review and meta-analysis. *Frontiers in Psychology*, *6*, Article 900. https://doi.org/10.3389/ fpsyg.2015.00900
- Bourdieu, P., & Passeron, J.-C. (1970). La reproduction : Éléments pour une théorie du système d'enseignement [The reproduction: Elements for a theory of the education system]. Éditions de Minuit.
- Bourdieu, P. (with Wacquant, L.). (2013). Symbolic capital and social classes. *Journal of Classical Sociology*, *13*(2), 292–302. https://doi. org/10.1177/1468795x12468736
- Bourdieu, P. (2000). Cultural reproduction and social reproduction. In R. Arum & I. Beattie (Eds.), *The structure of schooling: Readings in the sociology of education* (pp. 71–112). Mayfield Publishing.
- Lamont, M., & Lareau, A. (1988). Cultural capital: Allusions, gaps and glissandos in recent theoretical developments. *Sociological Theory*, 6(2), 153–168. https://doi.org/10.2307/202113
- 10. Lareau, A. (2011). Unequal childhoods: Class, race and family life (2nd ed. with a decade-later update). University of California Press.
- 11. Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006, September 1). Reducing the racial achievement gap: A social-psychological intervention. *Science*, 313(5791), 1307–1310. https:// doi.org/10.1126/science.1128317

- Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009, April 17). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science*, 324(5925), 400–403. https://doi. org/10.1126/science.1170769
- Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010, November 26). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, 330(6008), 1234–1237. https://doi.org/10.1126/ science.1195996
- Goyer, J. P., Garcia, J., Purdie-Vaughns, V., Binning, K. R., Cook, J. E., Reeves, S. L., Apfel, N., Taborsky-Barba, S., Sherman, D. K., & Cohen, G. L. (2017). Self-affirmation facilitates minority middle schoolers' progress along college trajectories. *Proceedings of the National Academy of Sciences, USA*, 114(29), 7594–7599. https://doi. org/10.1073/pnas.1617923114
- Cohen, G. L., & Sherman, D. K. (2014). The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology*, 65, 333–371. https://doi.org/10.1146/ annurev-psych-010213-115137
- 16. Levitt, S. D., List, J. A., Neckermann, S., & Sadoff, S. (2016). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy*, 8(4), 183–219. https://doi.org/10.1257/ pol.20130358
- Mayer, S. E., Kalil, A., Oreopoulos, P., & Gallegos, S. (2019). Using behavioral insights to increase parental engagement: The parents and children together intervention. *Journal of Human Resources*, 54(4), 900–925. https://doi. org/10.3368/jhr.54.4.0617.8835R
- Walton, G. M. (2014). The new science of wise psychological interventions. *Current Directions in Psychological Science*, 23(1), 73–82. https://doi. org/10.1177/0963721413512856
- Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research*, 81(2), 267–301. https://doi. org/10.3102/0034654311405999
- Piper, B., Dryden-Peterson, S., Chopra, V., Reddick, C., & Oyanga, A. (2020). Are refugee children learning? Early grade literacy in a refugee camp in Kenya. Journal on Education in Emergencies,

5(2), 71–107. https://doi.org/10.33682/ f1wr-yk6y

- 21. World Bank. (2018). World development report 2018: Learning to realize education's promise. https://doi. org/10.1596/978-1-4648-1096-1
- 22. Global Education Monitoring Report Team. (2019). Global education monitoring report youth report 2019: Migration, displacement, and education: Building bridges, not walls. UNESCO. https://unesdoc.unesco.org/ark:/48223/ pf0000366741
- 23. United Nations Development Programme Arab States, Nudge Lebanon, & B4Development. (2021). Applying behavioural science to support the prevention of violent extremism: Experiences and lessons learned. United Nations Development Programme. https://pveportal.org/ wp-content/uploads/2021/08/ undp-behavioural-science-supportthe-prevention-of-violent-extremism. pdf
- 24. Peters, E., Shoots-Reinhard, B., Tompkins, M. K., Schley, D., Meilleur, L., Sinayev, A., Tusler, M., Wagner, L., & Crocker, J. (2017). Improving numeracy through values affirmation enhances decision and STEM outcomes. *PLOS ONE*, *12*(7), Article e0180674. https:// doi.org/10.1371/journal.pone.0180674
- Borman, G. D., Grigg, J., & Hanselman, P. (2016). An effort to close achievement gaps at scale through self-affirmation. *Educational Evaluation and Policy Analysis*, 38(1), 21–42. https://doi. org/10.3102/0162373715581709
- 26. See, B. H., Morris, R., Gorard, S., θ Siddiqui, N. (2020). Writing about values: Addendum. Education Endowment Foundation. https://d2tic4wvo1iusb. cloudfront.net/documents/projects/ WAVE_Addendum_final.pdf
- 27. See, B. H., Morris, R., Gorard, S., & Siddiqui, N. (2018). Writing about values: Evaluation report and executive summary. Education Endowment Foundation. https://d2tic4wvo1iusb. cloudfront.net/documents/projects/ Writing_About_Values_1.pdf
- Wu, Z., Spreckelsen, T. F., & Cohen, G. L. (2021). A meta-analysis of the effect of values affirmation on academic achievement. *Journal of Social Issues*, 77(3), 702–750. https://doi.org/10.1111/ josi.12415
- Easterbrook, M. J., Harris, P. R., & Sherman, D. K. (2021). Self-affirmation theory in educational contexts. *Journal* of Social Issues, 77(3), 683–701. https:// doi.org/10.1111/josi.12459

- Borman, G. D. (2017). Advancing values affirmation as a scalable strategy for mitigating identity threats and narrowing national achievement gaps. Proceedings of the National Academy of Sciences, USA, 114(29), 7486–7488. https://doi.org/10.1073/ pnas.1708813114
- 31. Cazabat, C. (2019). *Twice invisible: Accounting for internally displaced children*. Internal Displacement Monitoring Centre. https://www. internal-displacement.org/ sites/default/files/publications/ documents/201911-twice-invisibleinternally-displaced-children.pdf
- 32. United Nations High Commissioner for Refugees. (2019). *Refugee education* 2030: A strategy for refugee inclusion. https://reliefweb.int/attachments/ abd47541-f405-3932-8ec1-9380c4b17df5/71213.pdf
- 33. Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811. https://doi. org/10.1037/0022-3514.69.5.797
- 34. Riciputi, S., & Erdal, K. (2017). The effect of stereotype threat on student-athlete math performance. *Psychology of Sport* and Exercise, 32, 54–57. https://doi. org/10.1016/j.psychsport.2017.06.003
- 35. Hess, T. M., Auman, C., Colcombe, S. J., & Rahhal, T. A. (2003). The impact of stereotype threat on age differences in memory performance. *The Journals of Gerontology: Series B*, 58(1), P3–P11. https://doi.org/10.1093/geronb/58.1.P3
- 36. Open Science Collaboration. (2015, August 28). Estimating the reproducibility of psychological science. *Science*, 349(6251), Article aac4716. https://doi.org/10.1126/science.aac4716
- Hall, C. C., Zhao, J., & Shafir, E. (2014). Self-affirmation among the poor: Cognitive and behavioral implications. *Psychological Science*, 25(2), 619–625. https://doi. org/10.1177/0956797613510949
- Altay, N., Kilicarslan-Toruner, E., & Sari, Ç. (2017). The effect of drawing and writing technique on the anxiety level of children undergoing cancer treatment. *European Journal of Oncology Nursing*, 28, 1–6. https://doi.org/10.1016/j. ejon.2017.02.007
- Piasai, K., Phumdoung,
 S., Wiroonpanich, W., &
 Chotsampancharoen, T. (2018). A randomized control trial of guidedimagination and drawing-storytelling

in children with cancer. *Pacific Rim International Journal of Nursing Research, 22*(4), 386–400.

- 40. Abbing, A., Ponstein, A., van Hooren, S., de Sonneville, L., Swaab, H., & Baars, E. (2018). The effectiveness of art therapy for anxiety in adults: A systematic review of randomised and non-randomised controlled trials. *PLOS ONE*, *13*(12), Article e0208716. https:// doi.org/10.1371/journal.pone.0208716
- Feldman, L., & Hart, P. S. (2018). Is there any hope? How climate change news imagery and text influence audience emotions and support for climate mitigation policies. *Risk Analysis*, 38(3), 585–602. https://doi.org/10.1111/ risa.12868
- 42. Powell, T. E., Boomgaarden, H. G., De Swert, K., & de Vreese, C. H. (2015). A clearer picture: The contribution of visuals and text to framing effects. *Journal of Communication*, 65(6), 997–1017. https://doi.org/10.1111/ jcom.12184
- Parrott, S., Hoewe, J., Fan, M., θ Huffman, K. (2019). Portrayals of immigrants and refugees in U.S. news media: Visual framing and its effect on emotions and attitudes. *Journal of Broadcasting θ Electronic Media*, 63(4), 677–697. https://doi.org/10.1080/08838 151.2019.1681860
- 44. Bruner, J. (1992). Another look at New Look 1. *American Psychologist*, 47(6), 780–783. https://doi. org/10.1037/0003-066X.47.6.780
- 45. Kitayama, S., Duffy, S., Kawamura, T., & Larsen, J. T. (2003). Perceiving an object and its context in different cultures: A cultural look at New Look. *Psychological Science*, *14*(3), 201–206. https://doi. org/10.1111/1467-9280.02432
- 46. Bruner, J., & Goodman, C. C. (1947). Value and need as organizing factors in perception. *The Journal of Abnormal* and Social Psychology, 42(1), 33–44. https://doi.org/10.1037/h0058484
- 47. Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. Annual Review of Psychology, 67, 415–437. https://doi.org/10.1146/ annurev-psych-073115-103235
- 48. Eschenbach, E. A., Virnoche, M., Cashman, E. M., Lord, S. M., & Camacho, M. M. (2014). Proven practices that can reduce stereotype threat in engineering education: A literature review. In 2014 IEEE Frontiers in Education Conference (FIE) Proceedings. Institute of Electrical and Electronics Engineers. https://doi. org/10.1109/FIE.2014.7044011

- 49. Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613–629. https:// doi.org/10.1037/0003-066X.52.6.613
- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. *Journal* of Experimental Social Psychology, 35(1), 29–46. https://doi.org/10.1006/ jesp.1998.1371
- 51. Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *British Journal of Educational Psychology*, *77*(2), 323–338. https://doi.org/10.1348/000709906X113662
- 52. Appel, M., Kronberger, N., & Aronson, J. (2011). Stereotype threat impairs ability building: Effects on test preparation among women in science and technology. *European Journal of Social Psychology*, 41(7), 904–913. https://doi. org/10.1002/ejsp.835
- 53. Borman, G. D., & Pyne, J. (2016). What if Coleman had known about stereotype threat? How social-psychological theory can help mitigate educational inequality. RSF: The Russell Sage Foundation Journal of the Social Sciences, 2(5), 164–185. https://doi. org/10.7758/RSF.2016.2.5.08
- 54. Hanselman, P., Bruch, S. K., Gamoran, A., & Borman, G. D. (2014). Threat in context: School moderation of the impact of social identity threat on racial/ ethnic achievement gaps. *Sociology of Education*, 87(2), 106–124. https://doi. org/10.1177/0038040714525970
- 55. Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues*, 57(1), 55–71. https://doi. org/10.1111/0022-4537.00201
- 56. Smith, J. L. (2004). Understanding the process of stereotype threat: A review of mediational variables and new performance goal directions. *Educational Psychology Review*, *16*(3), 177–206. https://doi.org/10.1023/ B:EDPR.0000034020.20317.89
- Keith, P. M., Byerly, C., Floerchinger, H., Pence, E., & Thornberg, E. (2006). Deficit and resilience perspectives on performance and campus comfort of adult students. *College Student Journal*, 40, 546–556.

- Sullivan, A. L., & Simonson, G. R. (2016). A systematic review of schoolbased social-emotional interventions for refugee and war-traumatized youth. *Review of Educational Research*, 86(2), 503–530. https://doi. org/10.3102/0034654315609419
- 59. Hamilton, R., & Moore, D. (Eds.). (2003). Educational interventions for refugee children: Theoretical perspectives and implementing best practice. Routledge. https://doi.org/10.4324/9780203687550
- Richmond, M., Robinson, C., & Sachs-Israel, M. (2008). The global literacy challenge: A profile of youth and adult literacy at the mid-point of the United Nations Literacy Decade 2003–2012 (Document ED.2008/WS/52). UNESCO. https://unesdoc.unesco.org/ark:/48223/ pf0000163170
- Burde, D., Kapit, A., Wahl, R. L., Guven, O., & Skarpeteig, M. I. (2017). Education in emergencies: A review of theory and research. *Review of Educational Research*, 87(3), 619–658. https://doi. org/10.3102/0034654316671594
- 62. United Nations High Commissioner for Refugees. (2022, December). *Lebanon* [Fact sheet]. https://www.unhcr.org/lb/ wp-content/uploads/sites/16/2023/01/ UNHCR-Lebanon-Operational-FactSheet-Year-end-2022_Final-rev..pdf
- 63. Ministry of Education and Higher Education. (2018, July). RACE II fact sheet. https://www.dropbox.com/s/ qqvih82odfnbu2y/MEHE_REC_Fact_ Sheet_July_2018.pdf
- 64. المدركة الرتيوي للبحوث واإلمناء. النشرة الإلحصائية: للعام الدراسي 2018–2017 [Statistical bulletin: School year 2017– 2018]. https://www.crdp.org/sites/ default/files/2019-11/2_0.pdf
- 65. Salloukh, B. F. (2017). The Syrian war: Spillover effects on Lebanon. *Middle East Policy*, 24(1), 62–78.
- 66. Gazit, O. (2021). What it means to (mis)trust: Forced migration, ontological (in)security, and the unrecognized political psychology of the Israeli-Lebanese conflict. *Political Psychology*, 42(3), 389–406. https://doi.org/10.1111/ pops.12703
- Helou, M., El-Hussein, M., Aciksari, K., Salio, F., Della Corte, F., von Schreeb, J., & Ragazzoni, L. (2022). Beirut explosion: The largest non-nuclear blast in history. *Disaster Medicine and Public Health Preparedness*, 16(5), 2200–2201. https:// doi.org/10.1017/dmp.2021.328

- 68. Baumann, H. (2019). The causes, nature, and effect of the current crisis of Lebanese capitalism. *Nationalism and Ethnic Politics*, *25*(1), 61–77. https://doi. org/10.1080/13537113.2019.1565178
- 69. McQueen, A., & Klein, W. M. P. (2006). Experimental manipulations of selfaffirmation: A systematic review. *Self and Identity*, *5*(4), 289–354. https://doi. org/10.1080/15298860600805325
- Makki, F., Mohanna, N., & Shephard, D. (2018, June 17). Visual values affirmation to increase student testing performance (RCT ID AEARCTR-0003081) [Preregistration]. AEA RCT Registry. https://doi.org/10.1257/rct.3081-1.0
- Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the statistical representation of the effects of educational interventions into more readily interpretable forms* (NCSER 2013-3000). Institute of Education Sciences National Center for Special Education Research. https://ies.ed.gov/ ncser/pubs/20133000/pdf/20133000. pdf
- 72. Williams, C. L., Hirschi, Q., Sublett, K. V., Hulleman, C. S., & Wilson, T. D. (2020). A brief social belonging intervention improves academic outcomes for minoritized high school students. *Motivation Science*, 6(4), 423–437. https://doi.org/10.1037/mot0000175
- Sarouphim, K. M., & Chartouny, M. (2017). Mathematics education in Lebanon: Gender differences in attitudes and achievement. *Educational Studies in Mathematics*, 94(1), 55–68. https://doi.org/10.1007/ s10649-016-9712-9
- 74. Gajderowicz, T. J., & Jakubowski, M. J. (2022). Jordan and Lebanon performance in international student assessments. World Bank. https:// openknowledge.worldbank.org/ handle/10986/37631
- 75. Critcher, C. R., & Dunning, D. (2015). Self-affirmations provide a broader perspective on self-threat. *Personality and Social Psychology Bulletin*, 41(1), 3–18. https://doi. org/10.1177/0146167214554956
- 76. Oh, J. S., & Fuligni, A. J. (2010). The role of heritage language development in the ethnic identity and family relationships of adolescents from immigrant backgrounds. *Social Development*, *19*(1), 202–220. https://doi. org/10.1111/j.1467-9507.2008.00530.x

- 77. Hourani, A. (1983). Arabic thought in the liberal age 1798–1939. Cambridge University Press. https://doi.org/10.1017/ CBO9780511801990
- 78. Smith, E. N., Rozek, C. S., Manke, K. J., Dweck, C. S., & Walton, G. M. (2021). Teacher- versus researcher-provided affirmation effects on students' task engagement and positive perceptions of teachers. *Journal of Social Issues*, 77(3), 751–768. https://doi.org/10.1111/ josi.12417
- 79. United Nations Development Programme Arab States, Nudge Lebanon, & B4Development. (2021). Applying behavioural science to support the prevention of violent extremism: Experiences and lessons learned. United Nations Development Programme. https://www.undp.org/publications/ applying-behavioural-science-supportprevention-violent-extremismexperiences-and-lessons-learned
- Dowell, N. M. M., McKay, T. A., θ Perrett, G. (2021). It's not *that* you said it, it's *how* you said it: Exploring the linguistic mechanisms underlying values affirmation interventions at scale. *AERA Open*, 7, 1–19. https://doi. org/10.1177/23328584211011611
- 81. M. Schwalbe, personal communication, January 31, 2022.



finding

Encouraging selfblinding in hiring

Sean Fath, Richard P. Larrick, & Jack B. Soll

abstract

One strategy for minimizing bias in hiring is *blinding*-purposefully limiting the information used when screening applicants to that which is directly relevant to the job and does not elicit bias based on race, gender, age, or other irrelevant characteristics. Blinding policies remain rare, however. An alternative to blinding policies is self-blinding, in which people performing hiring-related evaluations blind themselves to biasing information about applicants. Using a mock-hiring task, we tested ways to encourage selfblinding that take into consideration three variables likely to affect whether people self-blind: default effects on choices, people's inability to assess their susceptibility to bias, and people's tendency not to recognize the full range of information that can elicit that bias. Participants with hiring experience chose to receive or be blind to various pieces of information about applicants, some of which were potentially biasing. They selected potentially biasing information less often when asked to specify the applicant information they wanted to receive than when asked to specify the information they did not want to receive, when prescribing selections for other people than when making the selections for themselves, and when the information was obviously biasing than when it was less obviously so. On the basis of these findings, we propose a multipronged strategy that human resources leaders could use to enable and encourage hiring managers to self-blind when screening job applicants.

Fath, S., Larrick, R. P., & Soll, J. B. (2023). Encouraging self-blinding in hiring. *Behavioral Science & Policy*, 9(1), 45–57. https://doi.org/10.1353/XXXXX

n a study published in 2017, researchers analyzed roughly five years of pull requeststhat is, proposed changes to software projects-on the software development site GitHub to test whether requests made by women were evaluated differently than those made by men. The results were striking. When proposed changes came from software developers who were outsiders to a project (as opposed to project owners or known collaborators), project leaders were more likely to accept changes proposed by men than those proposed by women. However, this trend held only when the gender of the developer proposing the changes was identifiable. When project leaders were unable to discern the gender of the developer proposing the changes, they became more likely to accept proposals from women than from men.1

This example demonstrates the value of a policy of *blinding*, or purposefully limiting the availability of irrelevant information that could potentially bias an evaluation of a person's ideas, gualifications, or performance. Blinding policies increase objectivity in evaluations by preventing evaluators from receiving information that might bias their assessments. In the domain of work, a hiring manager who does not know the name of a job applicant-say, because the name has been stripped from the applicant's resume-cannot possibly use that name to make assumptions about the applicant's race, gender, or other attributes peripheral to job performance. Stereotypes about race and gender cannot then leak into assessments of other information, such as job credentials.²

Yet when it comes to making hiring decisions, blinding policies remain relatively rare. Although a handful of boutique firms, such as GapJumpers and Applied, have emerged to help companies perform blind initial screens of job applicants, we have found that few institutions choose to use such services or establish internal blinding policies for the hiring process. In a survey we reported on in 2021, we asked more than 800 human resources (HR) professionals—who averaged 14 years of experience in the field—about whether they had experience with blinding policies in the hiring process.³ We found that 81% of them had never worked at an organization that used blinding policies at any point during hiring. Moreover, 80% indicated they had never received training about blinding as a possible bias-reduction strategy.

Although these data are not representative of all U.S. organizations, they suggest that blinding policies and services are not commonly used in hiring. Some other alternative hiring practices, such as artificial intelligence–based screens of applicants, may be considered blind to the extent they are machine-driven, but these practices can still result in biased evaluations. For instance, automatically screening out applicants who have gaps in employment can affect women disproportionately, because women have employment gaps more often than men do.^{4,5} Further, machine-driven practices are often used in conjunction with nonblind human evaluations.^{6–8}

Blinding may be uncommon in institutional hiring in part because the jobs of most organizations vary widely in gualifications and duties. As a result, HR professionals may be concerned that uniform rules may not be appropriate in all cases. In addition, they may want to avoid limiting the autonomy of managers in making hiring decisions, as hiring managers tend to balk at initiatives such as diversity-fostering hiring policies that limit their latitude in decision-making.9 Yet without blinding, hiring decisions may be compromised by bias from information that is not directly related to job qualifications. Biasing information-such as a person's name, age, or appearance-is often either included in applicant materials¹⁰ or easily gathered from the internet.11,12

In this article, we ask, is it possible to encourage those making hiring-related decisions to selfblind—to choose on their own not to receive biasing information about applicants? Encouraging self-blinding during the initial screening of applicants would preserve hiring managers' autonomy as well as the flexibility needed to adapt the hiring process to particular jobs. For instance, an organization could introduce a checklist-based system by which a manager could pick which information to see or not to see when evaluating job candidates. Such a system represents a behavioral *nudge*—a gentle push to do something that does not limit autonomy or choices. It could prompt managers to avoid seeing biasing information without limiting their freedom, thereby reducing employment discrimination based on race, age, gender, or any other job-irrelevant characteristic.

To address our question, we explored the influence of three key factors on whether hiring managers performing an initial screen of applicants would blind themselves to biasing information about those applicants: the psychological pull of defaults, people's sense of their susceptibility to bias, and people's understanding of what information can lead to bias. We examined the effects of each of these factors in a mock hiring task, with the overall aim of determining the most effective design for a self-blinding process in organizations. Next, we discuss the science undergirding our exploration of these factors and our predictions about their effects.

Factors Influencing the Likelihood of Self-Blinding

Default Effects

Hiring decisions are typically structured such that hiring managers receive biasing information about job applicants by default. For instance, hiring managers often learn applicants' names at the beginning of the hiring process, and a name may provide information such as a person's race, gender, and social class. Biases about these social categories can then distort the way the manager evaluates the person's suitability for the job. To avoid this distortion, a hiring manager could choose to remain unaware of applicants' names, but that scenario is unlikely if managers get this information by default. The literature on default effects shows that decision-makers in many domains tend to accept defaults.

For instance, employees are more likely to participate in a retirement savings plan when their employer enrolls them by default, relative to when the default state is nonenrollment and participation requires employees to make an effort to sign up, or opt in, to the plan.¹³ People are more likely to be organ donors,¹⁴ undergo HIV screening,¹⁵ and get the flu vaccine¹⁶ when arrangements are made for them (forcing them to opt out to avoid participation) than when they must opt in to participate. (See the Supplemental Material for more information about default effects.)

Similarly, if hiring managers receive all information-including biasing information-about an applicant by default, they may be disinclined to depart from that default state to avoid receiving the biasing information. We therefore predicted that providing no information unless items were specifically requested (that is, unless managers opted in to receiving particular items) would result in managers being more likely to blind themselves to biasing information than would providing all information and requiring managers to opt out of seeing particular items. Put another way, we hypothesized that using an opt-in framework would be the optimal strategy for nudging hiring managers to blind themselves to biasing information.¹⁷

Perceived Susceptibility to Bias

Hiring managers' inclination to self-blind to biasing information about job applicants may also be shaped by their personal sense of susceptibility to bias in hiring decisions. Unfortunately, people are poor judges of their own propensity for bias. In social psychological research on self-perceived bias susceptibility, participants often judge themselves to be objective in their specific evaluations¹⁸ and general perceptions of the world¹⁹ and believe they are less susceptible to bias than others are.²⁰ This misperception may make hiring managers more likely to elect to see biasing information for themselves than they would be if they were making the choice for someone else. To test this proposition, we asked some of our participants to consider what choice they would make for others regarding whether they should see biasing information. We predicted that these participants would be more likely to avoid providing biasing information to others doing the screening than would participants instructed to make that choice for themselves. If that prediction proved correct, the finding

would indicate that people's misperception of their own susceptibility to bias at least partly affects whether they choose to look at biasing information. This misperception might be counteracted by asking managers to make a choice for someone else before choosing options for themselves.

Bias Transparency

The third factor that can influence whether hiring managers blind themselves to biasing information involves the nature of that information and whether managers recognize its potential to bias decisions. Some of the information that can bias decisions about job applicants may not be obviously biasing. For instance, an applicant's name may appear to be innocent background information even though it may indicate a person's gender and race, among other attributes. By contrast, explicit mention of a person's gender or race is transparently biasing. We tested how often participants chose to see transparently biasing versus nontransparently biasing information. We anticipated that participants would consider nontransparently biasing information to be less biasing than more overtly biasing information and thus would elect to see nontransparently biasing information more often than they would choose to see overtly biasing information. If so, this propensity would need to be considered when strategies nudging self-blinding are designed.

The Predictions in Brief

In a nutshell, we predicted that participants would be more likely to blind themselves to potentially biasing information (whether transparently or nontransparently biasing) when they had to opt in (specifically choosing what information to see) than when they had to opt out of receiving the information. We also predicted that participants would choose potentially biasing information for review less often when making the choice for others than when making it for themselves. Finally, we predicted that in any of those conditions, participants would elect to see transparently biasing information less often than they would elect to see nontransparently biasing information, even though both types could, in fact, bias their decisions.

We also tested whether self-blinding nudges might affect participants' interest in seeing information that is important for making a good decision. Strategies to encourage bias reduction by self-blinding should be adopted only if they do not markedly suppress hiring managers' inclination to receive useful information about job applicants-that is, information relevant to applicants' job qualifications. We did not expect self-blinding nudges to inhibit participants from electing to see information that is widely accepted to be diagnostic of job performance, because this information is not likely to be viewed as a source of bias. That is, we expected that participants would be just as likely to ask to see useful information regardless of the decision-making frame (opt in or opt out) or whether they were making the decision for themselves or for others.

Method

We recruited 800 participants with hiring experience to take part in our experiment, targeting about 100 participants for each of the eight study conditions we planned; we received 798 complete responses.²¹ The mean age of the participants was 39.82 years (SD = 12.03); 47.4% were women. Participants had an average of 19.16 years of work experience and estimated that they had made an average of 36.67 hiring decisions in their careers. They were all U.S. citizens and were recruited through an online platform (https://www.prolific.co/) that supplies research participants.

Participants completed a mock hiring task in which they screened applicants for a hypothetical position at their place of work to determine whom to advance to the interview stage. All participants received a checklist from which they could choose to see any of seven types of information available about applicants. Five of the seven items on the checklist represented useful information, which we define as information that is commonly accepted to be relevant for hiring decisions. We selected these items the job applicant's college, major, previous work experience, job-related skills, and references using a pool of sample applications for U.S. jobs posted online as a guide. The remaining two items on our checklist were those that we prejudged to be irrelevant to job performance and potentially biasing. Participants saw one of two sets of items, depending on their study condition. The first set of two items consisted of a job applicant's race and gender, which we deemed to be transparently biasing. The second set of two items consisted of a job applicant's picture and name, which we judged to be nontransparently biasing. All items were presented to participants in a randomized order. (All materials and data for our study are archived online at https://osf.io/2vthn/.)

To assess the effects of an opt-out or opt-in framework on self-blinding preferences, we randomly assigned participants to one of two sets of instructions: One told participants to tick the boxes next to the items they did not wish to receive (that is, to opt out of the default of receiving all the information), and the other told them to tick the boxes next to the items they wanted to receive (that is, to opt in to receiving specific information). To assess self-perceived susceptibility to bias, we randomly assigned the participants in the opt-out and opt-in conditions to either choose the information they wanted to receive if they were making the screening decision themselves or decide what information to provide to someone else doing the screening. Finally, we further divided those four groups, randomly assigning participants to use a checklist that included either the two transparently biasing items or the two nontransparently biasing items. For each of the resulting eight conditions (see Table 1), we tabulated the items participants chose to see.

To confirm that participants agreed with us on which items were useful versus biasing in relation to a hiring decision, we conducted a posttest using a separate group of 104 participants with hiring experience. The results generally supported our classifications of these items as useful (five items) or potentially biasing (two items). One exception was the name of the job applicant's college, which we had prejudged to be useful but posttest participants rated as slightly more biasing than useful. (See the Supplemental Material for details.) As a result, we did not use the name of the job applicant's college item in the analyses that follow.

Results

Overall, our hypotheses were supported. Participants were less likely to choose to see the biasing information (name, picture, gender, and race) when they were instructed to opt in to information they wanted to see (M = 22.3%) than when they had to opt out to exclude information they did not want to see (M = 32.1%, p < .001). Participants were also less likely to choose information that was potentially biasing when making a choice for others (M = 21.4%) than for themselves (M = 32.9%, p < .001). Finally, participants were less likely to elect to see biasing information when the possibility of bias was relatively transparent, as in a person's race or gender (M = 17%) than when it was

Table 1. The eight conditions studied, by mix of frame, person receiving the information, ϑ type of potentially biasing information available

Frame	Person receiving the information	Type of potentially biasing information
Opt out	Self	Transparently biasing
Opt out	Self	Nontransparently biasing
Opt out	Others	Transparently biasing
Opt out	Others	Nontransparently biasing
Opt in	Self Transparently biasing	
Opt in	Self	Nontransparently biasing
Opt in	Others	Transparently biasing
Opt in	Others	Nontransparently biasing

Note. The opt-out frame involves choosing what to exclude. The opt-in frame involves choosing what to include.

nontransparent, as in a person's picture or name (M = 37.3%, p < .001). These main effects are averages across all of the conditions. (See the Supplemental Material for more details of our analyses and results.)

Next, we compared the effects of the opt-in versus opt-out frames and self versus other decisions on choices to receive biasing versus useful applicant information. Across conditions, the vast majority of participants asked to see the useful information (M = 90.9%), whereas only about a quarter of the participants asked to see the biasing information (M = 27.2%, ps < .001).

When we looked at the effects of opting out or opting in specifically on the selection of transparently or nontransparently biasing information, we found that participants were less likely to choose to see the nontransparently biasing items (a picture or name) when they were in an opt-in condition in which they actively chose to see items (M = 31.3%) than when they were in an opt-out condition in which they excluded items from a list (M = 43.4%, p =.002). The same pattern held for the transparently biasing items (race and gender): A mean of 21.1% of the participants in an opt-out condition but a mean of only 12.8% of the participants in an opt-in condition selected the transparently biasing items (p = .015). However, for the useful items, the opt-out versus opt-in distinction had a much smaller effect, such that roughly 9 out of 10 participants chose to see the useful information regardless of the default frame. (The mean percentage of the opt-out conditions was 92.5%; that of the opt-in conditions was 89.4%, p = .026.)

Similarly, participants were less likely to choose the nontransparently biasing information for others (M = 30.2%) than for themselves (M =44.1%, p < .001), and the same was true for the transparently biasing items: A mean of 12.9% of participants in the other conditions chose these items compared with a mean of 21.3% in the self conditions (p = .015). In comparison, participants selected the useful items for themselves at roughly the same rate as they did for others, with about 9 out of 10 choosing the information in either case: A mean of 90.6% in the self conditions and a mean of 91.2% in the other conditions (p = .655).

The panels of Figure 1 break down the data further, showing the percentage of participants who chose for themselves (Panel A) or for others (Panel B) each of the useful and biasing items that were available, broken down by opt-out and opt-in conditions. Figure 2 shows the percentage of participants in each condition who chose useful information in aggregate (job skills, work experience, references, and college major) and biasing information in aggregate (name, picture, gender, and race).

Discussion

In our study, participants with hiring experience blinded themselves to information that was potentially biasing about mock job applicants more often when (a) they needed to opt in to see information about the applicants than when they had to opt out, (b) they were making a decision for someone else rather than for themselves, and (c) the biasing items were transparently biasing (such as the item identifying race) rather than more subtly biasing (such as the item providing a name). Next, we discuss ways that companies and other institutions may leverage these findings to encourage hiring managers and others making hiring-related decisions to blind themselves to potentially biasing information about job applicants.

Solutions

Leverage Default Effects. In our study, the opt-out scenario created a default in which participants would receive all the information on a checklist unless they opted out of some of it. In the opt-in scenario, the default was receiving no information. Research on default effects has predominantly demonstrated that people are more likely to adopt beneficial policies or behaviors under opt-out conditions than opt-in ones, such as when a person who does not want to be an organ donor has to opt out when obtaining a driver's license. In this study, however, the opt-in condition was more effective at minimizing the selection of biasing

Figure 1. Percentages of participants choosing for themselves ϑ for others whether to see applicant information



A. Participants choosing information for themselves

B. Participants choosing information for others



Note. When making a choice for themselves (A), participants generally asked to see biasing applicant information (four items at right) less often when they were given a list of possible items and asked to specify the ones they wanted to see (that is, when they were in an opt-in condition) than when they were instructed to specify which ones they did not want to see (that is, when they were in an opt-out condition). The same was true for participants who made choices for others (B). The same was not true for useful—that is, clearly job-relevant—information. In addition, whether choosing for themselves or for others, significantly fewer participants selected transparently biasing information (gender and race) than nontransparently biasing information (name and picture). Participants who chose information for others were less likely to seek biasing information than were those who chose information for themselves.

Figure 2. Percentages of participants choosing useful & biasing information, by condition



Note. Participants chose biasing information (name, picture, gender, and/or race) least often when they were in an opt-in condition and were choosing for others and most often when they were in an opt-out condition and choosing for themselves. The differences between conditions were attenuated for useful information.

information and did so without markedly diminishing interest in useful information relevant to job performance. In certain domains, such as hiring, some options (such as seeing job-related skills) are likely to be favored regardless of whether the option is provided by default.

So although most research on default effects underscores the effectiveness of interventions that allow people to make passive decisions,^{13,17} such as sticking with a desirable default, our work suggests that requiring active decision-making is best for nudging managers to self-blind to biasing information. In our study, participants who had to opt in to see information (that is, who made an active decision to look at each item) seemingly became more attentive to which items might bias their decisions and consequentially became less likely to select items providing biasing information. This takeaway is consistent with research demonstrating that inclusion frames (which require people to choose the best items from a broader list) foster more deliberative thinking than do exclusion frames (which require people to reject the worst items from a broader list).¹⁷ The findings are also consistent with research showing that when choices to receive biasing information are

driven by curiosity, curiosity-driven impulses can be reduced by using decision frames that cue deliberative reasoning.²²

Our results suggest that organizations could nudge hiring managers to selectively self-blind to biasing information by instituting a checklist system in which managers must pick the information they wish to see about applicants. An organization could have a dedicated employee create the checklist by itemizing the information available about applicants and then give that list to the hiring managers. Such a process may also be appealing to decision-makers, who tend to prefer opt-in to opt-out frames when making choices.²³

Circumvent the Bias Blind Spot. Our finding that participants selected biasing information for themselves more often than they did for others is consistent with other research showing that people perceive others to be more susceptible to bias than they themselves are.²⁰ In our study, this difference was attenuated when the information was patently useful and relevant. These results suggest that to encourage hiring managers to self-blind, organizations could train hiring managers to consider what information

they would give to someone else making a hiring decision before making that choice for themselves. Because hiring managers are likely to want their decisions to be consistent, they are likely to make the same choices for themselves as they did for others.²⁴ A training module encouraging hiring managers to "consider what information you would want someone else making this decision to have" could be included in organizations' antibias training and would likely offer benefits beyond encouraging selfblinding in hiring.

Boost Awareness of Hidden Bias. Our finding that participants selected nontransparently biasing information at a higher rate than they selected transparently biasing information is noteworthy because much of the biasing information available to hiring managers may be nontransparently biasing. For instance, applicants' names are commonly provided on applications, and photos are typically available on applicants' social media pages or personal websites, which are information sources managers often use.^{11,12,25} Although a name or a picture may be less obviously biasing than a person's noted gender or race, they are often just as biasing or more so. Both often convey race and gender, and a photo is likely to communicate additional biasing information such as attractiveness, age, and physical fitness. Many studies have documented hiring bias related to applicants' names.¹⁰

Yet our results show that many people are not aware of how biasing names and photos can be. Similarly, people might not realize that being aware of applicants' college graduation years may trigger biases related to age, that knowing applicants' hobbies could lead to biases related to social class or disability status,²⁶ and so on. Even information that is objectively nonbiasing and merely irrelevant to a hiring decision is a good candidate for blinding, because, at best, the inclusion of such information adds noise to evaluations.²⁷ Our results suggest that hiring managers would be more likely to selfblind to biasing information the more they are made aware-perhaps through continuing education-of the potentially biasing or at least noise-inducing content lurking within seemingly innocent information. To ensure that hiring managers get this education, organizations could require them to complete a training module on hidden sources of bias in hiring-related information.

Combat Belief in the Usefulness of Biasing Information. We understand that which applicant information is biasing versus relevant is likely to vary across industries and that even the information deemed biasing in our experiments may be directly relevant in some contexts. For instance, a photo is likely to be useful and relevant for a modeling job.

But hiring managers often believe that potentially biasing information is useful when it is not. In our study, for instance, participants may have sought to derive information about cultural fit from applicants' photos, as we have observed in other studies.²² It is also possible that some participants chose to view potentially biasing information because they wanted to favor applicants from marginalized groups. Although well-intentioned, this type of choice carries dangers, because removing unwanted bias from one's reasoning is difficult.² As an example, a well-intentioned hiring manager might seek out an applicant's photo to clarify their race or gender, with the goal of favoring applicants from marginalized groups. But in the process, that manager opens the door to biases related to age and attractiveness. Corporate training sessions on bias can help hiring managers understand that biasing information is typically more harmful than helpful, which, in turn, should increase their preference for self-blinding.

Other Considerations Related to Self-Blinding

Could self-blinding ever be counterproductive to combating discrimination? As we have noted, hiring managers may want to see information that will help them favor applicants from disadvantaged social groups. Another argument against blinding is that knowing applicants' social group status might provide important context for assessing their credentials. If, for example, a person had to overcome a lifetime of disadvantage or discrimination to gain those credentials, would it not be helpful to take that past into account? Doing so might help minimize the advantages of members of dominant groups whose qualifications might derive in part from privilege. It is generally for these reasons that the merits of "colorblind" policies are questioned.^{28,29}

However, the bulk of the field studies assessing the effects of using blind initial screens suggest that members of marginalized groups, such as women and ethnic minorities, are often more likely to reach the interview stage when initial screens are anonymized.³⁰⁻³² Moreover, multiple recent reviews and meta-analyses show that unblinded initial screens of applicants decrease the likelihood of members of marginalized groups receiving callbacks for interviews.^{10,33-36} These findings strongly indicate that a blinding process during applicant screening can, via a reduction in discrimination, help achieve an institution's goal of diversity in hiring.

Moreover, self-blinding in applicant screening may have carryover benefits in interviews. If the

same evaluator who performed an initial blind screen also interviews the selected job candidates, that evaluator is likely to continue to try to discount biasing demographic information to maintain a consistent strategy throughout the process.^{24,37}

Still, self-blinding is just one tool among many that may be used to achieve diversity goals in hiring. Self-blinding nudges should be used in tandem with other strategies, such as unblind targeted recruiting to increase the proportion of people from marginalized groups who apply for a job in the first place-for instance, by establishing talent development or pipeline programs at historically Black colleges and universities-and structured interviewing procedures that decrease the likelihood of bias against members of marginalized groups in face-to-face interviews. A multifaceted strategy is necessary to address bias in hiring decisions, and self-blinding is one important component of that strategy. (See the Supplemental Material for a fuller discussion of hiring procedures that promote diversity in hiring.)

In Brief: How Organizations Can Encourage Hiring Managers to Self-Blind

Human resources professionals can decrease the chances of bias in hiring practices in their organizations by enabling and encouraging hiring managers to blind themselves to biasing information about applicants during the initial screening process. Here are the actions we recommend, based on our research:

Appoint an intermediary to create a checklist. Assign a dedicated employee to read applications, categorize the information in them, and make a checklist for the hiring manager listing all the available types of information about applicants—for example, work experience, college major, references, name, and photograph.

Draft instructions for checklist use that encourage self-blinding.

- Ask hiring managers to first consider what information they would want another hiring manager to see. For example, "Imagine a situation in which someone else is tasked with hiring someone for an open position at your organization, and it is up to you to decide what information they incorporate into their decision." Then ask the manager to pick that information for themselves.
- Ask managers to pick the information they want to see rather than asking them to pick the information they do not want to see (that is, use opt-in rather than opt-out framing). For example, "Here are the pieces of information about applicants that are available. Please tick the box(es) next to the information you want to see."

Offer or require training for managers about hidden sources of bias and the importance of blinding themselves to biasing information. This training should help managers spot the potential for bias in information that does not necessarily seem like it would trigger biases, such as a person's name or photograph. Training could also help combat beliefs in the potential usefulness of biasing information.

Conclusion

Until blinding policies become commonplace in hiring, seemingly innocuous information about job applicants, such as their name, hobbies, or college graduation year, will continue to enable discrimination. In this article, we have discussed self-blinding, in which hiring managers choose on their own to avoid information about applicants that could bias or distort their evaluations, and we have identified three factors that could influence whether managers self-blind.

Our research suggests steps that organizational leaders could take to encourage hiring managers to self-blind when screening job applicants in the early stages of the hiring process (see the sidebar In Brief: How Organizations Can Encourage Hiring Managers to Self-Blind). We propose that organizations appoint a dedicated employee to itemize the available information about applicants. Evaluators could then be instructed to think about which types of information they would provide to a peer (because people tend to be stricter when choosing for others than for themselves) before opting in to receiving the information for themselves. Organizations should also provide training concerning the types of information that carry bias, myths about the usefulness of such information, and how to combat misperceptions of one's own susceptibility to bias. We believe that nudging self-blinding using the principles and guidelines outlined here will result in fairer and more accurate decisions in hiring.

author affiliation

Fath: Cornell University. Larrick and Soll: Duke University. Corresponding author's e-mail: sean. fath@cornell.edu.

supplemental material

- http://behavioralpolicy.org/journal
- Method & Analysis

references

- 1. Terrell, J., Kofink, A., Middleton, J., Rainear, C., Murphy-Hill, E., Parnin, C., & Stallings, J. (2017). Gender differences and bias in open source: Pull request acceptance of women versus men. *PeerJ Computer Science*, *3*, Article e111. https://doi.org/10.7717/peerj-cs.111
- Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*(1), 117–142. https://doi. org/10.1037/0033-2909.116.1.117
- 3. Fath, S., & Zhu, S. (2021). Preferences for, and familiarity with, blinding among HR practitioners. SSRN. https://doi. org/10.2139/ssrn.3768039
- 4. Ajunwa, I. (2019). The paradox of automation as anti-bias intervention. *Cardozo Law Review*, *41*(5), 1671– 1742. http://cardozolawreview.com/ the-paradox-of-automation-as-antibias-intervention/
- Ajunwa, I., Friedler, S., Scheidegger, C. E., & Venkatasubramanian, S. (2016). *Hiring by algorithm: Predicting and preventing disparate impact*. SSRN. http://sorelle.friedler.net/papers/SSRNid2746078.pdf
- Ajunwa, I. (2019, October 8). Beware of automated hiring. *The New York Times*. https://www.nytimes.com/2019/10/08/ opinion/ai-hiring-discrimination.html
- Teichman, D., Talley, E., Egidy, S., Engel, C., Gummadi, K., Hagel, K., Lewandowsky, S., MacCoun, R., Utz, S., & Zamir, E. (2021). Institutions promoting or countering deliberate ignorance. In R. Hertwig & C. Engel (Eds.), *Deliberate ignorance: Choosing not to know* (pp. 275–298). MIT Press. https://doi.org/10.7551/ mitpress/13757.003.0022
- 8. Lewis, N. (2018). Will AI remove hiring bias? Society for Human Resource Management. https://www.shrm. org/LearningAndCareer/learning/ educational-program-materials/ Documents/SHRM_Will%20AI%20 Remove%20Hiring%20Bias.pdf
- Dobbin, F., Schrage, D., & Kalev, A. (2015). Rage against the iron cage: The varied effects of bureaucratic personnel reforms on diversity. *American Sociological Review*, 80(5), 1014–1044. https://doi. org/10.1177/0003122415596416
- Bertrand, M., & Duflo, E. (2017). Field experiments on discrimination. In A. V. Banerjee & E. Duflo (Eds.), *Handbook* of economic field experiments (Vol. 1, pp. 309–393). Elsevier. https://doi. org/10.1016/bs.hefe.2016.08.004

- Acquisti, A., & Fong, C. (2020). An experiment in hiring discrimination via online social networks. *Management Science*, 66(3), 1005–1024. https://doi. org/10.1287/mnsc.2018.3269
- Bartoš, V., Bauer, M., Chytilová, J., & Matejka, F. (2016). Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review*, 106(6), 1437–1475. https://doi.org/10.1257/aer.20140571
- Madrian, B. C., & Shea, D. F. (2001). The power of suggestion: Inertia in 401(k) participation and savings behavior. *The Quarterly Journal of Economics*, 116(4), 1149–1187. https://doi. org/10.1162/003355301753265543
- 14. Johnson, E. J., & Goldstein, D. (2003, November 21). Do defaults save lives? *Science*, *302*(5649), 1338–1339. https:// doi.org/10.1126/science.1091721
- Montoy, J. C. C., Dow, W. H., & Kaplan, B. C. (2016). Patient choice in opt-in, active choice, and opt-out HIV screening: Randomized clinical trial. *The BMJ*, 352, Article h6895. https://doi. org/10.1136/bmj.h6895
- Chapman, G. B., Li, M., Colby, H., & Yoon, H. (2010). Opting in vs opting out of influenza vaccination. JAMA, 304(1), 43–44. https://doi.org/10.1001/ jama.2010.892
- Levin, I. P., Huneke, M. E., & Jasper, J. D. (2000). Information processing at successive stages of decision making: Need for cognition and inclusion– exclusion effects. *Organizational Behavior and Human Decision Processes*, *82*(2), 171–193. https://doi. org/10.1006/obhd.2000.2881
- Uhlmann, E. L., & Cohen, G. L. (2007). "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. Organizational Behavior and Human Decision Processes, 104(2), 207–223. https://doi.org/10.1016/j. obhdp.2007.07.001
- 19. Armor, D. A. (1999). *The illusion of objectivity: Bias in the belief in freedom from bias* [Unpublished doctoral dissertation]. University of California, Los Angeles.
- Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review*, 111(3), 781–799. https://doi. org/10.1037/0033-295X.111.3.781

- Gervais, W. M., Jewell, J. A., Najle, M. B., & Ng, B. K. (2015). A powerful nudge? Presenting calculable consequences of underpowered research shifts incentives toward adequately powered designs. *Social Psychological and Personality Science*, 6(7), 847–854. https://doi. org/10.1177/1948550615584199
- Fath, S. B., Larrick, R. P., & Soll, J. (2022). Blinding curiosity: Exploring preferences for "blinding" one's own judgment. Organizational Behavior and Human Decision Processes, 170, Article 104135. https://doi.org/10.1016/j. obhdp.2022.104135
- Goodman, J. K., & Reczek, R. W. (2021). Choosing what to choose from: Preference for inclusion over exclusion when constructing consideration sets from large choice sets. *Journal* of Behavioral Decision Making, 34(1), 85–98. https://doi.org/10.1002/ bdm.2199
- 24. Festinger, L. (1957). A theory of cognitive dissonance (Vol. 2). Stanford University Press.
- 25. Jobvite. (2014). 2014 Social Recruiting Survey [Slides]. https://www.jobvite. com/wp-content/uploads/2014/10/ Jobvite_SocialRecruiting_Survey2014. pdf
- 26. Rivera, L. A. (2012). Hiring as cultural matching: The case of elite professional service firms. *American Sociological Review*, 77(6), 999–1022. https://doi. org/10.1177/0003122412463213
- Hertwig, R., & Engel, C. (2016). Homo ignorans: Deliberately choosing not to know. Perspectives on Psychological Science, 11(3), 359–372. https://doi. org/10.1177/1745691616635594
- Markus, H. R., Steele, C. M., & Steele, D. M. (2000). Colorblindness as a barrier to inclusion: Assimilation and nonimmigrant minorities. *Daedalus*, 129(4), 233–259.
- Rattan, A., & Ambady, N. (2013). Diversity ideologies and intergroup relations: An examination of colorblindness and multiculturalism. *European Journal of Social Psychology*, 43(1), 12–21. https://doi.org/10.1002/ ejsp.1892
- Åslund, O., & Skans, O. N. (2012). Do anonymous job application procedures level the playing field? *ILR Review*, 65(1), 82–107. https://doi. org/10.1177/001979391206500105
- Bøg, M., & Kranendonk, E. (2011). Labor market discrimination of minorities? Yes, but not in job offers (No. 33332). Munich Personal RePEc Archive. https://mpra. ub.uni-muenchen.de/33332/1/MPRA_ paper_33332.pdf

- 32. Krause, A., Rinne, U., & Zimmermann, K. F. (2012). Anonymous job applications in Europe. *IZA Journal of European Labor Studies*, 1, Article 5. https://doi. org/10.1186/2193-9012-1-5
- Quillian, L., Heath, A., Pager, D., Midtbøen, A. H., Fleischmann, F., & Hexel, O. (2019). Do some countries discriminate more than others? Evidence from 97 field experiments of racial discrimination in hiring. Sociological Science, 6, 467–496. https://doi.org/10.15195/v6.a18
- 34. Quillian, L., Pager, D., Hexel, O., & Midtbøen, A. H. (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. Proceedings of the National Academy of Sciences, USA, 114(41), 10870–10875. https://doi.org/10.1073/ pnas.1706255114
- 35. Zschirnt, E., & Ruedin, D. (2016). Ethnic discrimination in hiring decisions: A meta-analysis of correspondence tests 1990–2015. Journal of Ethnic and Migration Studies, 42(7), 1115– 1134. https://doi.org/10.1080/1369 183X.2015.1133279
- 36. Pager, D., & Shepherd, H. (2008). The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annual Review of Sociology*, *34*, 181–209. https://doi.org/10.1146/annurev. soc.33.040406.131740
- Stone, J., & Fernandez, N. C. (2008). To practice what we preach: The use of hypocrisy and cognitive dissonance to motivate behavior change. Social and Personality Psychology Compass, 2(2), 1024–1051. https://doi. org/10.1111/j.1751-9004.2008.00088.x



Teamwork doesn't just happen: Policy recommendations from over half a century of team research

Gudela Grote and Steve W. J. Kozlowski

abstract

Teamwork has been at the core of human social organization for millennia and is essential for organizational productivity and innovation. Yet teamwork often is not as effective as it could be. Drawing on extensive research into the factors that enable teams to function well, this article offers policy recommendations for bolstering teamwork capabilities in society at large and in organizations. Our proposals call for teaching teamwork skills as part of the curricula in higher education and in lower grades in school, creating government and industry regulations designed to enhance teamwork, and designing jobs and organizational workflows in ways that prioritize and support teamwork.

Grote, G., & Kozlowski, S. W. J. (2023). Teamwork doesn't just happen: Policy recommendations from over half a century of team research. *Behavioral Science & Policy*, 9(1), 59–76. https://doi.org/10.1353/xxxxxxxx Teams of people working together for a common purpose have been a centerpiece of human social organization ever since our ancient ancestors first banded together to hunt game, raise families, and defend their communities.

-Steve W. J. Kozlowski & Daniel R. Ilgen¹

N April 11, 1970, three astronauts boarded Apollo 13, the spacecraft carrying out the United States' third mission to the moon. Two days into the flight, an oxygen tank exploded, causing extensive damage to the craft. Within three hours, the oxygen stores were gone and, along with them, the craft's ability to generate electrical power and operate its life support systems. A team of engineers back on Earth had to figure out how to get the crew home safely. Their seamless communication, determination, and adaptability, among other critical assets, led to a historic success: Despite the failed mission, teamwork saved the crew.

The rescue of the Apollo 13 crew shows the lifesaving power of excellent teamwork under the most demanding conditions.¹ On the flip side, poor teamwork can lead to disaster. It played a role in the 1986 explosions at the Chernobyl nuclear power plant in the former Soviet Union; the 1984 release of toxic chemicals in Bhopal, India; and the 2010 Deepwater Horizon explosion and massive oil spill in the Gulf of Mexico—all of which are legendary for their consequential losses of human life, extensive environmental damage, and substantial financial costs.

Acknowledging the drastic consequences of poor teamwork, civil aviation became the first industry to systematically promote a teamwork culture. Because of overwhelming evidence that many accidents are the result of aircrews' failure to collaborate well, it made teamwork training obligatory for all commercial pilots.^{2,3} (We describe the requirements in more detail in the Regulating for Teamwork section of this article.)

Teamwork is essential for the success of a large variety of industries and pursuits, in environments from the shop floor to the halls of

academia.^{4–6} In short, teamwork matters, and the push for organizational agility in response to increasing pressures to innovate has made effective collaboration in teams even more important.^{1,7,8} Indeed, technology giants such as Google have acknowledged teamwork's centrality and declared teamwork to be core to their success, and team-based methods that originated from managing software development have become cornerstones of organizational transformation across various industries.^{9,10}

We should note that although most people intuitively understand teamwork to be collaboration by a group to achieve goals important to an organization, investigators who study the topic also apply more formal definitions (see the sidebar What Is a Team? What Is Teamwork?).

Going forward, the ability to foster effective teamwork will become even more critical to organizations' ability to thrive, because at the same time reliance on teams is growing, teams themselves are becoming more complex. Artificial intelligence (AI) embodied in robots, other autonomous entities, and decision support systems are enabling flexible collaborations in which technology takes over certain tasks, supports decisions, and provides guidance.^{11,12} Good teamwork is essential to ensuring that these hybrid human-technology systems are effective. The global COVID-19 pandemic accelerated the evolution of digitally enabled teamwork. This metamorphic shift to virtual teaming requires good teamwork skills that transcend the lack of face-to-face contact. An increasingly diverse workplace also necessitates attention to teamwork. Although diversity offers a heterogeneity of views, experience, and ideas that can boost creativity, it also can create substantial challenges for collaboration.

What Is a Team? What Is Teamwork?

Investigators who study teams and teamwork have developed detailed definitions of the terms, such as those that follow.

In 1992, Eduardo Salas and his colleagues defined a team as "a distinguishable set of two or more people who interact dynamically, interdependently, and adaptively toward a common and valued goal/ objective/mission."^A

More recently, Steve W. J. Kozlowski and Daniel R. Ilgen expanded the definition to "(a) two or more individuals; (b) who interact socially (often face-to-face, but increasingly virtual); (c) possess one or more common goals; (d) are formed to perform organizationally relevant tasks; (e) exhibit interdependencies with respect to workflow, goals, and outcomes; (f) have a differentiated structure of roles and responsibilities; (g) and are embedded in an encompassing organizational system, with boundaries and linkages to the broader context and task environment."^B

Salas and his colleagues have also highlighted differences between *taskwork* and *teamwork*: "Taskwork involves the performance of specific tasks that team members need to complete in order to achieve team goals... teamwork focuses more on the shared behaviors (i.e., what team members *do*), attitudes (i.e., what team members *feel* or *believe*), and cognitions (i.e., what team members *think* or *know*) that are necessary for teams to accomplish these tasks."^c

A. From page 4 of Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Sweeney & E. Salas (Eds), *Teams: Their training and performance* (pp. 3–29). Ablex.
B. From page 79 of Kozlowski, S. W. J., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest*, 7(3), 77–124. https://doi.org/10.1111/j.1529-1006.2006.00030.x
C. From page 600 of Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L., & Lazzara, E. H. (2015). Understanding and improving teamwork in organizations: A scientifically based practical guide. *Human Resource Management*, *54*(4), 599–622. https://doi.org/10.1002/hrm.21628

Now is therefore a good time to build on insights from industries and organizations that have taken teamwork seriously and to scale up efforts to promote team effectiveness. Yet many organizations still focus primarily on building task-related technical skills while giving much less attention to teamwork skills such as solving problems collaboratively, resolving conflicts, and supporting one another, which are required to accomplish tasks.^{1,5}

We suspect that the ongoing inattention to strengthening teamwork in many organizations stems in part from a paradox: On the one hand, many organizational leaders still hold the view that teamwork happens easily without any extra effort by anyone; on the other hand, companies often have difficulty setting up programs that succeed in fostering teamwork.

In this article, we aim to alter the perception that teamwork processes happen effortlessly. A team of experts does not automatically make an expert team. In the next section, we point to key findings from an extensive body of research that has identified the core processes used by successful teams and the skills and capabilities that underlie those processes.^{1,13–18} In the sections that follow, we offer evidence-based recommendations for interventions that support effective teamwork. We have not attempted to be exhaustive in our recommendations, given the wide range of options to promote teamwork in organizations and beyond, nor do we provide a comprehensive literature review. Rather, we provide a concise overview and summarize the most salient evidence that undergirds our recommendations. Thus, we emphasize particularly relevant and impactful avenues for action based on key insights and strong evidence from team research.

We propose that teamwork skills be taught and assessed as part of school curricula from a young age; that, in the work arena, licensing requirements for individuals and organizations include teamwork training and the assessment of teamwork skills to increase individuals' and organizations' readiness to invest in improving teamwork; and that enterprises not only provide teamwork training but also specifically "Many organizations focus primarily on building taskrelated technical skills while giving much less attention to teamwork skills."

organize themselves in ways that enhance team effectiveness.

Good teamwork requires effort and training. Researchers know a lot about what makes teams effective, but society in general and organizations in particular need to find better ways to act on that knowledge. We hope that our recommendations stimulate such action.

The Science of Effective Teamwork

Teamwork has been studied for decades. The human relations movement in the 1930s and 1940s established that social factors such as group cohesion and recognition contribute to individuals' performance.¹⁹ And early studies on teamwork in coal mining showed that beyond satisfying social needs, teams are crucial for accomplishing complex and highly interdependent tasks.²⁰

Since then, the science of teamwork has evolved into a specialty. Teams are complicated entities. Although they are composed of individuals, they have collective properties that emerge from the individuals' interactions in the context of their task and organizational system. Thus, individuals are nested in teams, teams are nested in the broader organizational system, and the interconnections among these levels evolve dynamically over time.²¹ To address this complexity, research into what makes for an effective team has examined teamwork capabilities, or competencies, that are rooted in individuals but that lead to effective teamwork at the collective level. Interventions target both levels.

Several targeted¹³⁻¹⁵ and more comprehensive^{1,16-18} reviews conducted over the past two decades have compiled the extensive evidence identifying the core teamwork processes that support team effectiveness, the capabilities that underlie good teamwork, and key interventions that shape good teamwork. Of particular note, a comprehensive review by Steve W. J. Kozlowski and Daniel R. Ilgen amassed evidence from meta-analyses (which statistically combine data from multiple studies) showing that particular teamwork processes contribute to team effectiveness.1 These core teamwork processes are concisely summarized in Table 1. Kozlowski and Ilgen also highlighted key interventions with significant empirical support for enhancing the targeted teamwork processes; these methods are concisely summarized in Table 2. The sidebar Core Capabilities encapsulates the core capabilities that underlie good teamwork processes.²²⁻²⁵ These capabilities are the primary targets for team training and other interventions designed to improve team effectiveness. See Tables S1, S2, and S3 in the Supplemental Material for more detailed descriptions of key findings.

As Table 1 shows, substantial scientific evidence has identified three core teamwork processes that enable teams to be successful. Teams need to build a common basis for action through sharing knowledge, they need to continuously adapt their knowledge and actions to fit changing situational demands, and they need to keep team members motivated to contribute to shared team goals.

For example, investigations of the knowledge component have found that measures of information sharing and team cognition correspond to measures of the effectiveness of team performance and decision-making. By *information sharing*, we mean team members communicating information that everyone needs to know and that relates to specific expertise. By *team cognition*, we mean team members sharing the common, organized knowledge they need and also understanding who knows other key information. (The term *organized knowledge* refers to how various facts and concepts relate to one another.) When teams share information well,

Process	Aspect studied	What research examines	Key findings
Share knowledge widely	Information sharing	The extent to which common and specialized information is shared in a team	Information sharing is significantly related to team performance overall (ρ = .42). ^a
			When both shared information and unique information need to be combined for optimal decisions, team members tend to share the general information two standard deviations more than unique information. Teams that do this are eight times less likely to make the correct decision than are teams with full access to all the information. ^b
	Team cognition	The extent to which all team members have shared, organized information needed by the team as a whole as well as distributed knowledge connected by a shared understanding of who knows what specific knowledge	Measures of team cognition correlate well with measures of team behavioral processes ($\rho = .43$), motivational states ($\rho = .43$), and performance ($\rho = .38$). Distributed knowledge affects team performance ($\rho = .44$) more than shared information does ($\rho = .32$). ^c
Adapt readily to changing circumstances and related behaviors	Team goals	How the nature of the team's goals affects attentiveness, strategy, and effort	Setting difficult and specific group-level goals boosts group performance one standard deviation more than no goals or low-level goals. ^d
			Group-level goals improve group performance ($d = 0.56 \pm 0.19$, $k = 49$) more when they are difficult and specific than when they are easy to achieve and general or vague ($d = 0.80 \pm 0.35$, $k = 23$). ^e
	Collective behavior	Which team-oriented behaviors contribute to a team's ability to adapt	Relevant behaviors are communicating (ρ = .22), coordinating (ρ = .30), adjusting plans and strategies in response to disruptions (ρ = .41), learning as a group (ρ = .27), and formulating a specific plan for the team to reach a goal (ρ = .24). ^f Team cognition is positively related to team adaptive performance (ρ = .19). ^f
Maintain high levels of motivation	Team cohesion	Whether strong bonds among members influence team performance	Group cohesion is significantly related to group performance (33 effect sizes; $\rho = .317$). ⁹ Cohesion is more strongly linked to a team's performance when the team's task is highly interdependent—requiring a
			lot of team coordination ($\rho = .464$)—than when the need for coordination is low ($\rho = .206$). ^h
	Team efficacy	Whether a shared belief that the team can collectively overcome challenges affects performance	Team efficacy is significantly associated with team performance (ρ = .41), and more so when the task is highly interdependent (low, ρ = .09; high, ρ = .47). ⁱ

Table 1. Processes of effective teams & related findings from meta-analyses

Note. See Table S1 in the Supplemental Material for additional information about the meta-analyses. See note A for a discussion of the statistics used in the tables in this article.

a. Mesmer-Magnus, J. R., & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. Journal of Applied Psychology, 94(2), 535–546. https://doi.org/10.1037/a0013773

b. Lu, L., Yuan, C., & McLeod, P. L. (2012). Twenty-five years of hidden profiles in group decision making: A meta-analysis. Personality and Social Psychology Review, 16(1), 54–75. https://doi.org/10.1177/1088868311417243

c. DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: A meta-analysis. Journal of Applied Psychology, 95(1), 32–53. https://doi.org/10.1037/a0017328

d. O'Leary-Kelly, A. M., Martocchio, J. J., & Frink, D. D. (1994). A review of the influence of group goals on group performance. Academy of Management Journal, 37(5), 1285–1301. https://doi.org/10.2307/256673

e. Kleingeld, A., van Mierlo, H., & Arends, L. (2011). The effect of goal setting on group performance: A meta-analysis. Journal of Applied Psychology, 96(6), 1289–1304. https://doi.org/10.1037/a0024315

f. Christian, J. S., Christian, M. S., Pearsall, M. J., & Long, E. C. (2017). Team adaptation in context: An integrated conceptual model and meta-analytic review. *Organizational Behavior and Human Decision Processes*, 140, 62–89. https://doi.org/10.1016/j.obhdp.2017.01.003

g. Gully, S. M., Devine, D. J., & Whitney, D. J. (1995). A meta-analysis of cohesion and performance: Effects of levels of analysis and task interdependence. Small Group Research, 26(4), 497–520. https://doi.org/10.1177/1046496495264003

h. Beal, D. J., Cohen, R. R., Burke, M. J., & McLendon, C. L. (2003). Cohesion and performance in groups: A meta-analytic clarification of construct relations. Journal of Applied Psychology, 88(6), 989–1004. https://doi.org/10.1037/0021-9010.88.6.989

i. Gully, S. M., Incalcaterra, K. A., Joshi, A., & Beaubien, J. M. (2002). A meta-analysis of team-efficacy, potency, and performance: Interdependence and level of analysis as moderators of observed relationships. *Journal of Applied Psychology*, 87(5), 819–832. https://doi.org/10.1037/0021-9010.87.5.819

Intervention	Description	Key findings
Team training	Educating students or employees on how to work on a team through lectures; exercises; and, for intact teams, simulations of real-world situations and challenges	Team training improves team cognition (ρ = .42), affect (ρ = .35), process (ρ = .44), and performance (ρ = .39). ^a
		Leadership training improves employees' attitudes toward the training (δ = .63; 95% CI [0.12, 1.15]), leadership skills and knowledge (δ = .73; 95% CI [0.62, 0.85]), and leadership performance in the workplace (δ = .82; 95% CI [0.58, 1.06]). It also yields benefits to the organization, such as increased profits or reduced employee turnover (δ = .72; 95% CI [0.60, 0.84]). ^b
		Team training of various types—but principally workshop exercises and simulations of specific tasks—has medium to large effects on team behaviors and large effects on team performance across contexts ranging from aviation to academia, $d(0.13) = 0.683$, 95% CI [0.43, 0.94], $Z = 5.23$, $p < .001$; $Q(38) = 660.7$, $l^2 = 94.2$. ^c
		After-action reviews (debriefings) significantly improve team attitudes, cognition, processes, and performance (sample weighted mean $d = 0.79$, $SD = 0.83$, 95% CI [0.63, 0.95]). ^d
Work design	Distributing tasks and workflow among team members in a way that is motivating, delineating where these tasks overlap or depend on one another, and providing the resources a team needs to perform those tasks	A work design that gives team members autonomy is associated with improved job performance, both objectively ($\rho = .17$) and subjectively ($\rho = .23$). Other aspects of work design are also associated with improved performance. These include the degree to which people can complete a whole piece of work ($\rho = .17$), the extent to which the work affects others' lives ($\rho = .23$), and the amount of feedback employees receive ($\rho = .20$). ^e
Climate	Having a shared understanding that values teamwork and ensures that team members understand their collective mission, the reasoning behind it, their roles, and their priorities	A teamwork climate increases team members' commitment to the organization and their life satisfaction, thereby improving job performance and psychological well-being and reducing signs of disinterest in the work. ^f Meta-analytical findings indicate that perceptions of the work climate are significantly related to attitudes about work, psychological well-being, motivation, and performance. ^g

Table 2. Methods for improving teamwork in organizations & related findings from meta-analyses

Note. See Table S2 in the Supplemental Material for additional information about the meta-analyses. See note A for a discussion of the statistics used in the tables in this article.

a. Salas, E., DiazGranados, D., Klein, C., Burke, C. S., Stagl, K. C., Goodwin, G. F., & Halpin, S. M. (2008). Does team training improve team performance? A metaanalysis. *Human Factors*, *50*(6), 903–933. https://doi.org/10.1518/001872008X375009

b. Lacerenza, C. N., Reyes, D. L., Marlow, S. L., Joseph, D. L., & Salas, E. (2017). Leadership training design, delivery, and implementation: A meta-analysis. Journal of Applied Psychology, 102(12), 1686–1718. https://doi.org/10.1037/apl0000241

c. McEwan, D., Ruissen, G. R., Eys, M. A., Zumbo, B. D., & Beauchamp, M. R. (2017). The effectiveness of teamwork training on teamwork behaviors and team performance: A systematic review and meta-analysis of controlled interventions. *PLOS ONE, 12*(1), Article e0169604. https://doi.org/10.1371/journal. pone.0169604

d. Keiser, N. L., & Arthur, W., Jr. (2021). A meta-analysis of the effectiveness of the after-action review (or debrief) and factors that influence its effectiveness. Journal of Applied Psychology, 106(7), 1007–1032. https://doi.org/10.1037/apl0000821

e. Humphrey, S. E., Nahrgang, J. D., & Morgeson, F. P. (2007). Integrating motivational, social, and contextual work design features: A meta-analytic summary and theoretical extension of the work design literature. *Journal of Applied Psychology*, 92(5), 1332–1356. https://doi.org/10.1037/0021-9010.92.5.1332

f. Carr, J. Z., Schmidt, A. M., Ford, J. K., & DeShon, R. P. (2003). Climate perceptions matter: A meta-analytic path analysis relating molar climate, cognitive and affective states, and individual level work outcomes. *Journal of Applied Psychology*, 88(4), 605–619. https://doi.org/10.1037/0021-9010.88.4.605

g. Parker, C. P., Baltes, B. B., Young, S. A., Huff, J. W., Altmann, R. A., LaCost, H. A., & Roberts, J. E. (2003). Relationships between psychological climate perceptions and work outcomes: A meta-analytic review. *Journal of Organizational Behavior*, 24(4), 389–416. https://doi.org/10.1002/job.198

Core Capabilities

1. Develop team strategies and goals.

- 2. Coordinate interdependent tasks.
- 3. Monitor task progress (goals).
- 4. Monitor team processes.
- 5. Provide feedback and support.
- 6. Promote collaborative problem-solving.
- 7. Foster team cohesion and endurance.
- 8. Address and resolve conflict.

Note. For teams to work effectively, team members need to master the eight skills listed above, which underlie the three team processes described in Table 1. See Table S3 in the Supplemental Material for selected studies that support the value of these capabilities.

grasp the tasks, agree on what is important, and understand who knows what specific information, they can avoid wasting time on tangential activities, miscommunication, and meandering searches for information, and it facilitates decision-making.

Research into adaptation looks at how the nature of a team's goals affects strategy and effort and at which team behaviors contribute to a team's ability to adapt. Among the findings are that team goals that are more difficult and specific are more strongly related to team performance than are more general team goals. In addition, team cognition is associated with team adaptive performance, and a range of specific teamwork behaviors are related to team adaptation, including communication, coordination, and plan formation.

With respect to motivation, investigators have found that it is associated with a team's social cohesiveness and the shared belief that the team can overcome difficulties. These features are more critical to success when a team's task requires a lot of interdependence—that is, when the extent to which each person's ability to contribute to the goal depends on other people's actions. For example, a soccer team's prowess depends on coordinated action among team members, whereas a track team's success relies far more on individual performances. Studies of teamwork spanning some 75 years have delineated eight core competencies that underly the three essential team processes that have been identified. As summarized in the sidebar Core Capabilities, team members must be able to work together to develop strategies and goals for the team, coordinate task execution, monitor progress toward reaching goals and how well team processes are working, provide feedback and support, promote problem-solving, foster cohesion and endurance, and manage conflict.^{22–25} These competencies are the primary targets of training to improve teamwork.

The interventions that shape teamwork processes and hence team effectiveness involve team training, work design, and climate. Regarding team training, a large body of evidence indicates that both training aimed at team members and training aimed at team leaders have a substantial influence on improving team cognition, teamwork processes, and team performance (see Table 2 for evidence and references). Work design involves the distribution of tasks in a team, the interdependencies among tasks, and the resources and demands related to those tasks. A good work design ensures team members can apply their knowledge and skills to the team's goals and remain engaged in the work; work design has been shown to affect job performance. Climate refers to the shared assumptions and norms of the team. An effective team climate is one in which teamwork is valued and collaboration is the norm. It is one in which team members understand their mission, the reasoning behind it, their roles and their priorities, and what is rewarded and punished by management. Climate is related to psychological well-being, motivation, and performance.

With respect to implementing interventions, team training is flexible and broadly applicable, and it can be implemented in a variety of ways, as we shall discuss. Work design is under the control of organizational management and thus is specific to particular team task contexts. Similarly, team climate is substantially influenced by team leaders and the broader organizational system, making it context specific.

"Team training is a particularly potent intervention for improving team effectiveness regardless of the setting."

Team training is a particularly potent intervention for improving team effectiveness regardless of the setting.²⁶ The type of training provided in schools or the workplace will depend on whether the skill is generic across teams of all sorts or specific to a given team.²² Capabilities such as problem-solving and conflict resolution have generic aspects that can be taught to individuals in school, work, or any of a variety of settings. In contrast, the best ways to coordinate team activities and develop strategies for meeting team goals usually need to vary by context and so are more appropriately taught to intact work teams. Some generic capabilities-problem-solving, for example-may have team- or task-specific aspects that are also best addressed to intact teams. Thus, as a general strategy, schools and universities should offer courses that address the generic aspects of core teamwork capabilities, and workplaces should include training for the more specific aspects.

We now turn attention to policy recommendations for improving teamwork by members of society in general. We then address specific actions for teamwork in work environments.

Educating Students for Teamwork

The basic skills and abilities needed in teamwork, such as communication, collaborative problem-solving, and conflict negotiation, should be taught as part of standard school curricula. To date, teachers in primary and secondary schools mainly emphasize generic social skills that help children get along in the classroom.²⁷ For teamwork skills, they rarely offer the systematic instruction that they apply for other subjects, such as languages or science.⁵ As a result, students lack teamwork skills. To rectify this, the Organisation for Economic Co-operation and Development (OECD) has taken initial steps toward bringing formal teamwork training to schools. It established the assessment of basic social skills in schools as part of a program to better understand and support children's and juveniles' social and emotional skills development.²⁸ In another important initiative, the OECD evaluated high school students' ability to collaborate with others to solve a problem.²⁹ It found that only 8% of students across all 38 OECD countries showed a high degree of competence at skills such as being aware of group dynamics, ensuring compliance with agreed-upon roles, and resolving conflicts.

Similarly, Arthur C. Graesser and his colleagues have outlined several possible methods for teaching collaborative problem-solving and teamwork skills in schools.⁵ Those methods include conducting case-based analyses of realworld teamwork scenarios, as well as reflecting on the practice of working in a team. Although these methods have not been well studied yet, we recommend that early teamwork instruction combine teaching of psychological processes in teams with practical skills training, such as role-plays on managing interpersonal conflict or in-class demonstrations of the challenges of sharing information and making decisions in groups. Teachers could address these skills in the context of group projects, which would then be graded not only on the guality of the end product but also on the extent to which students worked together effectively.

Assessment of teamwork skills could include giving exams that test knowledge of the skills, strategies, and concepts needed for successful teamwork and grading students on practical exercises that give them a chance to display these skills. In its assessment of collaborative problem solving, the OECD established four skill levels based on students' ability to both solve complex problems and do so collaboratively. Educators could use these levels to establish a starting point for teaching teamwork skills to high school students.⁵
Learning teamwork skills in schools would be expected to increase students' later value in the labor market, given that recent economic analyses show that high-paying jobs increasingly require good social skills.³⁰ From an economic perspective, this is explained by the fact that social skills reduce coordination costs in highly specialized work processes, which means that social skills foster the good teamwork needed to accomplish highly interdependent tasks.

The importance of teamwork skills has received more attention at colleges and universities, probably because surveys of employers and university alumni consistently show that graduates are ill prepared for the social demands of their jobs.^{31,32} However, most colleges do not offer formal courses on teamwork. Instead, instructors teach teamwork skills informally by assigning students to group projects and assisting them in managing these projects.^{33,34} Arguably, teamwork skills cannot be taught fully in a classic classroom lecture. But relegating these skills to informal learning signals that they are less important than technical skills, making students less motivated to learn them.

Recognizing these problems, some universities have begun to offer formal courses on teamwork.^{35–38} An early example is provided by Gilad Chen and his colleagues, who have described in detail an elaborate course they developed at George Mason University in Virginia called The Psychology of Working in Groups and Teams. The course followed a framework developed by Michael A. Stevens and Michael J. Campion.³⁹ Instructors emphasized the core competencies of conflict resolution, collaborative problem-solving, communication, goal setting, performance management, and planning and task coordination.35 The course combined classroom lectures on teamwork with in-class exercises and simulations of real-world team situations at separate assessment centers. Students were evaluated by both in-class exams and their performance in assessment-center exercises. In evaluating the effectiveness of the approach, the researchers found that students in the course significantly outperformed control group students on a teamwork competencies test designed for the course. The control group students either had not had any teamwork instruction or had participated in the assessment center exercises but not the classroom lectures.

Universities should develop and routinely offer such teamwork courses. Courses could focus on particular teamwork situations, such as multidisciplinary research collaborations, virtual teams, or culturally diverse teams. They would improve individuals' competencies both as members of such teams and as team leaders. The National Academy of Sciences has created an impressive tool kit called Enhancing the Effectiveness of Team Science that is geared to scientific staff from administrators to graduate students and provides guidance on creating, supporting, and leading scientific teams.⁶ Universities can use such tools not only to support research teams with appropriate resources and policies but also to prepare students for managing the challenges of working in teams of all types.⁴⁰

Regulating for Teamwork

Government and industry regulations are powerful levers for change. Accordingly, we propose to include teamwork skills in professional and organizational licensing as a way to increase the awareness of their importance and the readiness to act on that awareness.

Civil aviation offers an excellent model. It has long been recognized that 60%-80% of aircraft accidents are the result of human error, with a substantial proportion of those errors caused by communication, coordination, or collaboration issues-that is, teamwork failures.² Correspondingly, commercial airlines are required to establish teamwork training programs to obtain a license to operate. This teamwork training, known in aviation as crew resource management training, is obligatory for all commercial pilots and is increasingly also required for flight attendants and air traffic controllers.^{2,3} Crew resource management training is built around so-called nontechnical skills, or notechs, focusing on cooperation, leadership, situation awareness, and decision-making.41 As part of the training, a range of specific behaviors must be taught and assessed (see Table 3). During

Competency	Competency description	Behavioral indicator
Communication	Demonstrates effective oral, non-verbal and written communications, in normal and non-normal situations.	Ensures the recipient is ready and able to receive the information
		Selects appropriately what, when, how and with whom to communicate
		Conveys messages clearly, accurately and concisely
		Confirms that the recipient correctly understands important information
		Listens actively and demonstrates understanding when receiving information
		Asks relevant and effective questions
		Adheres to standard radiotelephone phraseology and procedures
		Accurately reads and interprets required company and flight documentation
		Accurately reads, interprets, constructs and responds to datalink messages in English
		Completes accurate reports as required by operating procedures
		Correctly interprets non-verbal communication
		Uses eye contact, body movement and gestures that are consistent with and support verbal messages
Leadership and Teamwork	Demonstrates effective leadership and teamworking.	Understands and agrees with the crew's roles and objectives.
		Creates an atmosphere of open communication and encourages team participation
		Uses initiative and gives directions when required
		Admits mistakes and takes responsibility
		Anticipates and responds appropriately to other crew members' needs
		Carries out instructions when directed
		Communicates relevant concerns and intentions
		Gives and receives feedback constructively
		Confidently intervenes when important for safety
		Demonstrates empathy and shows respect and tolerance for other people
		Engages others in planning and allocates activities fairly and appropriately according to abilities
		Addresses and resolves conflicts and disagreements in a constructive manner
		Projects self-control in all situations
Problem Solving and Decision Making	Accurately identifies risks and resolves problems. Uses the appropriate decision-making processes.	Seeks accurate and adequate information from appropriate sources
		Identifies and verifies what and why things have gone wrong
		Employ(s) proper problem-solving strategies
		Perseveres in working through problems without reducing safety
		Uses appropriate and timely decision-making processes
		Sets priorities appropriately
		Identifies and considers options effectively
		Monitors, reviews, and adapts decisions as required
		Identifies and manages risks effectively
		Improvises when faced with unforeseeable circumstances to achieve the safest outcome

Table 3. Teamwork competencies required of pilots & criteria used to evaluate the pilots

Note. The criteria are listed in the Behavioral Indicator column. From *The Manual of Evidence-Based Training* (Appendix 1), by International Civil Aviation Organization, 2013. Copyright 2013 by the International Civil Aviation Organization.

this training, which is conducted in highly sophisticated flight simulators, cockpit crews, sometimes along with cabin crews, are exposed to critical situations—say, an engine failure, low fuel, bad weather, or some combination of problems—that the group has to resolve as a team. Instructors assess and debrief the crews on the teamwork skills, such as soliciting advice or providing emotional support, that the trainees demonstrated during the exercise. The trainees also receive formal classroom training to learn and reinforce these skills.

In the United States, fatal aircraft accidents have continuously decreased since the U.S. Federal Aviation Administration mandated crew resource management training for commercial airline flight crews.⁴² As will be seen next, other high-risk industries such as health care and nuclear power have followed suit.^{43,44}

Because of the growing awareness that medical errors and patient safety are substantially affected by teamwork, efforts comparable to those in civil aviation are beginning to take root in health care.⁴⁵⁻⁴⁷ In medicine, teamwork errors exact high costs in human life.^{46,47} Indeed, medical errors are the third leading cause of death in the United States; they may account for more than 250,000 deaths per year.⁴⁸ As in aviation, most of those human errors have their roots in poor teamwork.^{49,50}

A number of U.S. hospitals have deployed a validated program to improve medical teamwork called Team Strategies and Tools to Enhance Performance and Patient Safety (TeamSTEPPS).^{51,52} TeamSTEPPS emerged from a collaboration by the Agency for Healthcare Research and Quality and the U.S. Department of Defense. It is a freely available patient safety tool kit that targets four core teamwork competencies: leadership, communication, situation monitoring, and mutual support. In addition, a randomized controlled trial, the gold standard in medical research, demonstrated that training hospital emergency room physicians to be better team leaders led to better team leadership behavior and better patient care outcomes.53 To date, however, licensing for medical personnel does not require team training, so regulatory action is needed to make this training more widespread in the health care industry.

Many other high-risk industries where human life is at stake (such as nuclear power and railways) endeavor to improve teamwork skills through training. However, licensing of personnel as well as the license to operate for the respective organizations rarely depend on teamwork training. In industries with lower risk profiles and therefore less public attention and regulatory pressure, teamwork skills are even lower on regulators' agendas. That can be modified through policy action, by pushing for requirements in licensing procedures for professions and organizations across industries, and through more awareness of the importance of teamwork in professional associations. For instance, inadequacies in judicial counseling and decision-making have been attributed to the practice of relying on single lawyers. By convening teams of clients, judges, lawyers, and subject-matter experts at various stages of the judicial process and offering teamwork training to legal professionals, judicial counseling could be made more effective.⁵⁴ Similarly, problem-solving teams of schoolteachers could better support student learning and reduce inappropriate referrals to special education. Forming effective teams in schools would require team training for teachers and administrative support for the teams, among other changes that licensing requirements would encourage.55

As our comments imply, in addition to requiring training in teamwork skills, regulators can promote effective teamwork by establishing requirements that change work practices. In some industries, the license to operate depends on the ability of organizations to demonstrate that they create working conditions, norms, and values that are conducive to good teamwork. The catastrophic accident at Chernobyl in 1986 led international and national regulators to develop programs to instill a safety culture in the nuclear power industry.^{56,57} Team training is an important part of these programs—as a way of improving team processes such as communication, collaboration, and leadership-as is work design, which can ensure that teams have needed resources and personnel and that

"Both individuals and intact teams need training on how to work effectively as a team."

they distribute responsibilities effectively.⁵⁸⁻⁶⁰ Nuclear installations can benefit from guidelines, recommendations, and training materials designed to enhance teamwork provided by their own professional associations and by national and international regulators. They are also inspected and assessed on a regular basis to monitor progress toward establishing appropriate work practices. Similar inspection programs have been set up by regulators in aviation for commercial airlines and air traffic management providers.^{61,62} In health care, safety climate or culture is at the core of many organizational change programs.⁶³

In the financial services industry, the financial crash of 2008 spurred regulatory efforts to promote good work practices under the heading of "ethical culture."⁶⁴ Ethical culture regulations do not yet address teamwork specifically, but they should. Regulators could borrow practices from industries such as nuclear power and civil aviation that have made teamwork training part of a new safety culture.

Introducing team training, work design, and climate supportive of teamwork through regulatory action requires a participatory process and a tailored approach. For each profession and industry, core teamwork skills and methods for their assessment need to be defined. The eight core teamwork capabilities described earlier, such as collaborative problem-solving, coordination, and conflict management, are a good place to start, but the skills may need to be prioritized and assessed differently across industries. (For a discussion related to health care, see the article by Asela M. Olupeliyawa and his colleagues in the reference list.⁶⁵) Good team leadership may look quite different in a research and development team at a drug company, for example, than in a team of firefighters or tax lawyers. Task complexity, employee qualification, automation, and external relationships, to

name but a few factors, all need to be taken into account to promote effective change.⁶⁶

Once standards are in place, regulators need to be mindful that assessing work practices and climate is different from assessing technical installations and processes. Inspectors and auditors from regulatory bodies typically have engineering and science backgrounds aligned with the industries they regulate.67,68 These inspectors may lack the skills required to evaluate social, team, and organizational processes and may need to be trained in those skills. In addition, regulatory agencies will need to hire staff with social science backgrounds to ensure proper assessments as well as adequate feedback and support for the executives in charge of implementing the new licensing requirements within an organization.

Organizing for Teamwork

Educators can help students learn the necessary skills to collaborate in teams, and regulatory requirements and oversight can elevate the importance of good teamwork in organizations. In the final analysis, however, organizations have to bring all of these elements together in daily routines of good teamwork.⁶⁹ Across industries, organizations should take steps to capitalize on the power of teamwork and not just count on the efforts of top-performing individuals.⁷⁰ Organizational leaders should invest in team training, work design, and climate. (See Table 2 for key findings from research related to these actions, and see the sidebar Consequences of Poor Teamwork in the Supplemental Material for examples of the effects of poor teamwork in different fields.)

We have already described multiple examples of team training. We now discuss some of the research that offers broader insight into what constitutes effective training.^{26,71,72} The research shows, for instance, that both individuals and intact teams need training on how to work effectively as a team. This training can take the form of educators or management professionals teaching teamwork skills in classroom settings. Alternatively, experts might coach specific teams of, say, health care professionals or nuclear engineers through simulations of challenging or crisis situations.

Research also shows that training based on an analysis of the needs of a specific situation works best. For instance, in homogenous teams, it is important to increase awareness of complacency and groupthink, whereas in heterogenous teams, building shared mental models (organized information held collectively among a team) and a common language are more relevant.

Moreover, training should be designed so that teams develop and use the specific competencies and skills needed in their industry. For instance, a training program developed in a large teaching hospital focused on improving communication by exposing medical teams to situations that required team members to speak up during a simulated anesthesia delivery—for example, to urge a lead anesthetist to do a tracheotomy during a scenario in which a patient proved difficult to intubate.⁷³ In the evaluation of the training, the authors found that the training was particularly effective in getting team members to speak up when the debriefing emphasized assertiveness across hierarchies.

With respect to work design, it has been shown to strongly affect both individual and team performance and hence is a powerful lever for team effectiveness.^{74,75} Good design helps team members to apply their knowledge and skills to the team's tasks and remain engaged in the work. In particular, individual and team tasks should be designed in a way that allows for high levels of autonomy. That is, as organizations rightfully and increasingly rely on teams to coordinate interdependent tasks, adapt to continuously changing internal and external demands, and innovate to keep the organization ahead of its competitors, organizational leaders should make sure to give those teams sufficient control and resources.⁷⁶ In addition, work processes should be arranged so that employees feel their team's work is important and so that they receive feedback about their performance.77

Concerning climate, the aim is to develop shared norms and values that help teams understand

their roles and have trust in their colleagues and leaders. This often requires supportive organizational structures and policies as well as good leadership. Team leaders not only have an impact on the day-to-day functioning of teams but also play a crucial role in instigating change in team processes and outcomes.^{78,79} Efforts to effect change in organizations should therefore combine the introduction of new procedures with training that specifically prepares team leaders for their role in the changeover.

For example, in 2008, experts from Johns Hopkins University and the Michigan Health and Hospital Association included the training of team leaders in hospital intensive care units in a study aimed at improving teamwork to enhance patient safety-for instance, by controlling infections more effectively.63,80 They trained team leaders (one doctor and one nurse) at each of 67 hospitals (103 intensive care units) in best practices for infection control. They established a daily goals sheet to improve communication among clinicians and a safety program geared toward promoting a safety climate, among other measures. The intervention improved the staff's ratings of the team safety climate and substantially reduced catheter-related bloodstream infections at participating hospitals.

Conclusion

The sidebar Fostering Teamwork as a Society: Three Avenues of Action summarizes our recommended interventions. In following our recommendations, educators, regulators, and leaders at organizations need to also be mindful of changing conditions for teamwork. Changes in work practices, such as the shift to virtual teamwork during the pandemic, and technological advances require new skill sets and team arrangements. For successful human-AI teams, it will be crucial to embed what is known about effective teamwork into the design of these new systems. On the flip side, artificial agents can be programmed to assess, coach, and shape effective teamwork interactions on the fly.81 The earlier that teamwork knowledge and skills become part of policy action, the more likely it will be that these new technological opportunities will be exploited responsibly. We hope that our review and recommendations provide a compelling rationale and realistic call for such action.

The changes we call for, which build on more than half a century of team research, would affect the work of science itself. Universities would not only educate students in teamwork but also develop new programs to advance teamwork skills and establish a research culture based on multidisciplinary collaboration. In that way, academia would be better able to fulfill its promise to bring socially valuable innovation into a world of grand challenges that require large-scale collaborative efforts or, as it has been called, "team science."6,82,83 To establish this teamwork culture, universities would need to bolster and expand cross-disciplinary programs, and academia as a whole should reward faculty for collaborative efforts instead

of placing multidisciplinary research on the sidelines. For instance, research that integrates knowledge from several disciplines tends to be published in lower ranked journals and receives fewer citations than research in a single discipline.⁸⁴ In an encouraging move, funding agencies such as the National Institutes of Health and the National Science Foundation have begun to push for team science though calls for grant applications that require that multidisciplinary teams conduct the research. Agencies have begun to complement this effort with explicit guidance on how to set up, train, and support multidisciplinary research teams.⁸⁵

Acknowledging that teamwork skills need to complement technical expertise is fundamental. The psychological and behavioral sciences, which provide the knowledge base for good teamwork, are often considered soft because

Fostering Teamwork as a Society: Three Avenues of Action

- 1. Educating for teamwork: Teamwork skills should be part of the regular curriculum in K–12 schools and universities.
 - Elementary and middle school: Include role-plays on managing interpersonal conflict and in-class demonstrations of challenges of group information sharing and decision-making. Practice and evaluate skills in the context of group projects.
 - High school: Build on efforts by the Organisation for Economic Co-operation and Development to teach and assess social competencies and collaborative problem-solving.^A
 - University: Offer more courses that combine classroom teaching and formal assessment of teamwork skills with practice of those skills in simulated team situations. Use tool kits developed by the National Academy of Sciences to foster teamwork skills among scientists.^B
- 2. Regulating for teamwork: Licensing requirements for professions and organizations should include training and testing of teamwork skills.
 - Build on and adapt existing instruments from civil aviation for teamwork-related training and assessment requirements.
 - Build on and adapt regulatory requirements from the nuclear industry regarding work design and climate for teamwork effectiveness.
 - Train inspectors and auditors from regulatory bodies in the assessment of social, team, and organizational processes.
- 3. Organizing for teamwork: Organizations should increase their investment in interventions aimed at improving teamwork.
 - Design individual and team tasks so that team members have sufficient autonomy and adequate resources for self-management in the team.
 - Train whole teams on teamwork skills in the context of their organization.
 - Promote organizational and leadership development to build a climate of trust and support.

A. Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest*, *19*(2), 59–92. https://doi.org/10.1177/1529100618808244
 B. National Research Council. (2015). *Enhancing the effectiveness of team science*. The National Academies Press. https://doi.org/10.17226/19007

their variables cannot always be measured objectively and experimental designs to isolate causal relationships are difficult to implement due to practical and ethical concerns. Yet behavioral research has amassed sound evidence that organizations, be they firms or universities, should capitalize on to build the effective teamwork needed to succeed in the long run. Teamwork is at the core of modern society and should be nurtured with care and respect.

endnote

A. Editor's note to nonscientists: The p value of a statistical test is the probability of obtaining a result equal to or more extreme than would be observed merely by chance, assuming there are no true differences between the groups under study (this assumption is referred to as the null *hypothesis*). Researchers traditionally view *p* < .05 as the threshold of statistical significance, with lower values indicating a stronger basis for rejecting the null hypothesis. In addition to statistical significance, researchers consider the size of the observed effects, using such measures as Cohen's d or Cohen's h. Cohen's d or h values of 0.2, 0.5, and 0.8 typically indicate small, medium, and large effect sizes, respectively. Standard deviation (SD) is a measure of the amount of variation in a set of sample values. Approximately two-thirds of the observations fall between one standard deviation below the mean and one standard deviation above the mean. Standard error (SE) uses standard deviation to determine how precisely one has estimated a true population value from a sample. For instance, if one took enough samples from a population, the sample mean ± 1 standard error would contain the true population mean around two-thirds of the time. A 95% confidence interval (CI) for a given metric indicates that in 95% of random samples from a given population, the measured value will fall within the stated interval.

With regard to other data relating to the meta-analyses summarized in the tables with this

article, ρ (rho) indicates the strength of an association on a scale from -1.00 to +1.00, where 0 indicates no association and -1.00 or 1.00 indicates a perfect negative or positive association, respectively; *k* is the number of studies or distinct samples included in an analysis; δ is a form of Cohen's *d* that has been corrected for unreliability in the criterion; *d*(*SE*) is a sample-weighted standard error; *Z* is a measure of the statistical significance of the *d* value; *Q* is an estimate of the variability of effect sizes across studies; *df* in *Q*(*df*) indicates the degrees of freedom for the *Q* value; and *l*² estimates the proportion of the observed variance that reflects variance in true effect sizes.

author note

Steve W. J. Kozlowski gratefully acknowledges the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) for support from Grant W911NF2210005, awarded to a project for which Kozlowski was the principal investigator and that, in part, contributed to this work. The views, opinions, and findings contained in this article are solely those of the authors, and one should not construe them as an official ARI or Department of Defense position, policy, or decision, unless so designated by other documentation.

author affiliation

Grote: ETH Zürich. Kozlowski: University of South Florida. Corresponding author's e-mail: ggrote@ethz.ch.

supplemental material

- http://behavioralpolicy.org/journal
- Additional Tables & Sidebar

references

- Kozlowski, S. W. J., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest*, 7(3), 77–124. https://doi. org/10.1111/j.1529-1006.2006.00030.x
- Helmreich, R. L. (1997). Managing human error in aviation. *Scientific American*, 276(5), 62–67. https://doi. org/10.1038/scientificamerican0597-62
- Salas, E., Rozell, D., Driskell, J. D., & Mullen, B. (1999). The effect of team building on performance: An integration. Small Group Research, 30(3), 309–329. https://doi. org/10.1177/104649649903000303
- Fiore, S. M. (2008). Interdisciplinarity as teamwork: How the science of teams can inform team science. *Small Group Research*, 39(3), 251–277. https://doi. org/10.1177/1046496408317797
- Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest*, 19(2), 59–92. https://doi. org/10.1177/1529100618808244
- 6. National Research Council. (2015). Enhancing the effectiveness of team science. The National Academies Press. https://doi.org/10.17226/19007
- Jiang, K., Lepak, D. P., Hu, J., & Baer, J. C. (2012). How does human resource management influence organizational outcomes? A metaanalytic investigation of mediating mechanisms. *Academy of Management Journal*, 55(6), 1264–1294. https://doi. org/10.5465/amj.2011.0088
- Teece, D., Peteraf, M., & Leih, S. (2016). Dynamic capabilities and organizational agility: Risk, uncertainty, and strategy in the innovation economy. *California Management Review*, 58(4), 13–35. https://doi.org/10.1525/ cmr.2016.58.4.13
- 9. Robertson, B. J. (2015). *Holacracy: The new management system for a rapidly changing world*. Henry Holt and Company.
- 10. Romme, G. (2017). Management as a science-based profession: A grand societal challenge. *Management Research Review*, 40(1), 5–9. https://doi. org/10.1108/MRR-10-2016-0225
- 11. Abbass, H. A. (2019). Social integration of artificial intelligence: Functions, automation allocation logic and humanautonomy trust. *Cognitive Computation*, *11*(2), 159–171. https://doi.org/10.1007/ s12559-018-9619-0

- Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., & Horvitz, E. (2019). Beyond accuracy: The role of mental models in human-Al team performance. Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, 7(1), 2–11. https:// doi.org/10.1609/hcomp.v7i1.5285
- Gully S. (2000). Work teams research. In M. M. Beyerlein (Ed.), Work teams: Past, present and future (pp. 25–44). Kluwer Academic Publishers.
- 14. Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: From input-processoutput models to IMOI models. *Annual Review of Psychology*, 56, 517–543. https://doi.org/10.1146/annurev. psych.56.091103.070250
- Mathieu, J. E., Gallagher, P. T., Domingo, M. A., & Klock, E. A. (2019). Embracing complexity: Reviewing the past decade of team effectiveness research. Annual Review of Organizational Psychology and Organizational Behavior, 6, 17–46. https://doi.org/10.1146/ annurev-orgpsych-012218-015106
- Kozlowski, S. W. J., & Bell, B. S. (2003). Work groups and teams in organizations. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), Handbook of psychology: Vol. XII. Industrial/ organizational psychology (pp. 333–375). John Wiley & Sons. https:// doi.org/10.1002/0471264385.wei1214
- Kozlowski, S. W. J., & Bell, B. S. (2013). Work groups and teams in organizations. In N. Schmitt & S. Highhouse (Eds.), Handbook of psychology: Vol. 12. Industrial and organizational psychology (2nd ed., pp. 412–469). John Wiley & Sons. https:// doi.org/10.1002/9781118133880. hop212017
- Salas, E., Stagl, K. C., & Burke, C. S. (2004). 25 years of team effectiveness in organizations: Research themes and emerging needs. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 19, pp. 47–91). Wiley. https://doi. org/10.1002/0470013311.CH2
- 19. Mayo, E. (1945). *The social problems of an industrial civilization*. Harvard University.
- 20. Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the longwall method of coal-getting: An examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system. *Human*

Relations, 4(1), 3–38. https://doi. org/10.1177/001872675100400101

- Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions (pp. 3–90). Jossey-Bass.
- Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining team competencies and establishing team training requirements. In R. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 333–380). Jossey-Bass.
- Fleishman, E. A., & Zaccaro, S. J. (1992). Toward a taxonomy of team performance functions. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training* and performance (pp. 31–56). Ablex.
- LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, *61*(2), 273–307. https://doi. org/10.1111/j.1744-6570.2008.00114.x
- Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. Academy of Management Review, 26(3), 356–376. https://doi. org/10.5465/amr.2001.4845785
- Bisbey, T. M., Reyes, D. L., Traylor, A. M., & Salas, E. (2019). Teams of psychologists helping teams: The evolution of the science of team training. *American Psychologist*, 74(3), 278–289. https://doi.org/10.1037/ amp0000419
- 27. Bierman, K. L. (2020). The Fast Track Friendship Group program. In D. Nangle, C. Erdley, & R. Schwartz-Mette (Eds.), Social skills across the life span: Theory, assessment, and intervention (pp. 181–199). Academic Press. https://doi.org/10.1016/ B978-0-12-817752-5.00009-3
- 28. Kankaraš, M., & Suarez-Alvarez, J. (2019). Assessment framework of the OECD Study on Social and Emotional Skills (OECD Educational Working Papers No. 207). Organisation for Economic Co-operation and Development. https://doi. org/10.1787/5007adef-en

- Organisation for Economic Co-operation and Development. (2017). PISA 2015 results: Vol. V. Collaborative problem solving. https:// doi.org/10.1787/9789264285521-en
- 30. Deming, D. J. (2017). The growing importance of social skills in the labor market. *Quarterly Journal of Economics*, *132*(4), 1593–1640. https:// doi.org/10.1093/qje/qjx022
- Reyes, D. L., Dinh, J., Lacerenza, C. N., Marlow, S. L., Joseph, D. L., & Salas, E. (2019). The state of higher education leadership development program evaluation: A meta-analysis, critical review, and recommendations. *The Leadership Quarterly*, *30*(5), Article 101311. https://doi.org/10.1016/j. leaqua.2019.101311
- 32. Borrego, M., Karlin, J., McNair, L. D., & Beddoes, K. (2013). Team effectiveness theory from industrial and organizational psychology applied to engineering student project teams: A research review. *Journal of Engineering Education*, 102(4), 472–512. https://doi. org/10.1002/jee.20023
- Hughes, R. L., & Jones, S. K. (2011). Developing and assessing college student teamwork skills. *New Directions* for Institutional Research, 2011(149), 53–64. https://doi.org/10.1002/ir.380
- Neves, J. S., & Nakhai, B. (2016). A model for developing and assessing students' teamwork competencies. *Research in Higher Education Journal*, 31, 1–22. http://www.aabri.com/ manuscripts/162484.pdf
- Chen, G., Donahue, L. M., & Klimoski, R. J. (2004). Training undergraduates to work in organizational teams. Academy of Management Learning & Education, 3(1), 27–40. https://doi.org/10.5465/ amle.2004.12436817
- 36. Hobson, C. J., Strupeck, D., Griffin, A., Szostek, J., & Rominger, A. S. (2014). Teaching MBA students teamwork and team leadership skills: An empirical evaluation of a classroom educational program. American Journal of Business Education, 7(3), 191–212. https://doi. org/10.19030/ajbe.v7i3.8629
- Rapp, T. L., & Mathieu, J. E. (2007). Evaluating an individually selfadministered generic teamwork skills training program across time and levels. *Small Group Research*, 38(4), 532–555. https://doi. org/10.1177/1046496407300479
- Rennick, C., Hurst, A., Grove, J., θ Mohamed, S. (2018, October 11–12). Developing and assessing teamwork

skills in undergraduate students. In Educating for the future: Learning outcomes and experiential learning symposium. Toronto, Ontario, Canada. https://oucqa.ca/wp-content/ uploads/2018/10/2018-Program-Guide-FINAL.pdf

- Stevens, M. J., & Campion, M. A. (1994). The knowledge, skill, and ability requirements for teamwork: Implications for human resource management. *Journal of Management*, 20(2), 503–530. https://doi. org/10.1177/014920639402000210
- 40. Vogel, A. L., Hall, K. L., Fiore, S. M., Klein, J. T., Bennett, L. M., Gadlin, H., Stokols, D., Nebeling, L. C., Wuchty, S., Patrick, K., Spotts, E. L., Pohl, C., Riley, W. T., & Falk-Krzesinski, H. J. (2013). The team science toolkit: Enhancing research collaboration through online knowledge sharing. *American Journal of Preventive Medicine*, 45(6), 787–789. https://doi.org/10.1016/j. amepre.2013.09.001
- Avermaete, J. A G., van. (1998). NOTECHS: Non-technical skill evaluation in JAR-FCL (NLR-TP-1998-518). National Aerospace Laboratory, the Netherlands. http://hdl.handle. net/10921/1234
- 42. Aviation Safety Network. (2019). *By period*. Flight Safety Foundation. https:// aviation-safety.net/statistics/period/ stats.php
- 43. Weaver, S. J., Dy, S. M., & Rosen, M. A. (2014). Team-training in healthcare: A narrative synthesis of the literature. *BMJ Quality & Safety*, 23(5), 359–372. https:// doi.org/10.1136/bmjgs-2013-001848
- 44. International Atomic Energy Agency. (1996). Nuclear power plant personnel training and its evaluation (Technical Report Series No. 380). https:// www.iaea.org/publications/5714/ nuclear-power-plant-personneltraining-and-its-evaluation
- 45. Hughes, A. M., Gregory, M. E., Joseph, D. L., Sonesh, S. C., Marlow, S. L., Lacerenza, C. N., Benishek, L. E., King, H. B., & Salas, E. (2016). Saving lives: A meta-analysis of team training in healthcare. *Journal of Applied Psychology*, *101*(9), 1266–1304. https:// doi.org/10.1037/apl0000120
- 46. James, J. T. (2013). A new, evidencebased estimate of patient harms associated with hospital care. *Journal of Patient Safety*, 9(3), 122–128. https://doi. org/10.1097/PTS.0b013e3182948a69
- 47. Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (Eds.). (2000). *To err is*

human: Building a safer health system. National Academies Press. https://doi. org/10.17226/9728

- 48. Makary, M. A., & Daniel, M. (2016). Medical error—The third leading cause of death in the US. *The BMJ*, 353(8056), Article i2139. https://doi.org/10.1136/ bmj.i2139
- 49. The Joint Commission. (2016). *America's hospitals: Improving quality and safety* (2016 annual report). https://www.new-media-release. com/jointcommission/2016_annual_ report/2016-annual-report.pdf
- 50. Tomlinson, R., & Wakeling, S. (with Rockall, T.). (2019). *Learning from invited reviews: 2019 full report*. Royal College of Surgeons. https://invitedreviews. rcseng.ac.uk/full-report
- Health Research & Educational Trust. (2015). Improving patient safety culture through teamwork and communication: TeamSTEPPS[®]. https://www.aha. org/system/files/2018-01/2015_ teamstepps_FINAL.pdf
- Alonso, A., Baker, D. P., Holtzman, A., Day, R., King, H., Toomey, L., & Salas, E. (2006). Reducing medical error in the military health system: How can team training help? *Human Resource Management Review*, *16*(3), 396–415. https://doi.org/10.1016/j. hrmr.2006.05.006
- Fernandez, R., Rosenman, E. D., Olenick, J., Misisco, A., Brolliar, S. M., Chipman, A. K., Vrablick, M. C., Kalynych, C., Arbabi, S., Nichol, G., Grand, J., Kozlowski, S. W. J., & Chao, G. T. (2020). Simulation-based team leadership training improves team leadership during actual trauma resuscitations: A randomized controlled trial. *Critical Care Medicine*, 48(1), 73–82. https://doi.org/10.1097/ CCM.0000000000004077
- 54. Mosten, F. S., & Traum, L. (2018). Interdisciplinary teamwork in family law practice. *Family Court Review*, 56(3), 437–460. https://doi.org/10.1111/ fcre.12360
- 55. Rosenfield, S., Newell, M., Zwolski, S., Jr., & Benishek, L. E. (2018). Evaluating problem-solving teams in K–12 schools: Do they work? *American Psychologist*, 73(4), 407–419. https://doi.org/10.1037/ amp0000254
- 56. International Nuclear Safety Advisory Group. (1991). Safety culture (Safety Series No. 75-INSAG-4). International Atomic Energy Agency. https:// www-pub.iaea.org/MTCD/Publications/ PDF/Pub882_web.pdf

- 57. United States Nuclear Regulatory Commission. (2011). Safety culture policy statement. https://www.nrc.gov/ about-nrc/safety-culture/sc-policystatement.html
- Mearns, K., Kirwan, B., Reader, T. W., Jackson, J., Kennedy, R., & Gordon, R. (2013). Development of a methodology for understanding and enhancing safety culture in air traffic management. *Safety Science*, *53*, 123–133. https://doi. org/10.1016/j.ssci.2012.09.001
- Nieva, V. F., & Sorra, J. (2003). Safety culture assessment: A tool for improving patient safety in healthcare organizations. *BMJ Quality & Safety*, 12(Suppl. 2), ii17–ii23. https://doi. org/10.1136/qhc.12.suppl_2.ii17
- 60. Schöbel, M., Klostermann, A., Lassalle, R., Beck, J., & Manzey, D. (2017). Digging deeper! Insights from a multi-method assessment of safety culture in nuclear power plants based on Schein's culture model. *Safety Science*, 95, 38–49. https://doi.org/10.1016/j.ssci.2017.01.012
- International Civil Aviation Organization. (2006). Safety management manual (SMM) (Doc. 9859). https://tinyurl. com/4fnsn856
- 62. International Air Transportation Association. (2019). *Creating a positive safety culture* [White paper]. https:// go.updates.iata.org/safety-culture
- Pronovost, P. J., Berenholtz, S. M., Goeschel, C., Thom, I., Watson, S. R., Holzmueller, C. G., Lyon, J. S., Lubomski, L. H., Thompson, D. A., Needham, D., Hyzy, R., Welsh, R., Roth, G., Bander, J., Morlock, L., & Sexton, B. (2008). Improving patient safety in intensive care units in Michigan. *Journal* of Critical Care, 23(2), 207–221. https:// doi.org/10.1016/j.jcrc.2007.09.002
- 64. Treviño, L. K., Haidt, J., & Filabi, A. E. (2017). Regulating for ethical culture. Behavioral Science & Policy, 3(2), 57–72. https://doi.org/10.1353/bsp.2017.0013
- 65. Olupeliyawa, A. M., Hughes, C., θ Balasooriya, C. D. (2009). A review of the literature on teamwork competencies in healthcare practice and training: Implications for undergraduate medical education. South-East Asian Journal of Medical Education, 3(2), 61–72. https:// doi.org/10.4038/seajme.v3i2.454
- 66. Grote, G. (2012). Safety management in different high-risk domains—All the same? Safety Science, 50(10), 1983–1992. https://doi.org/10.1016/j.ssci.2011.07.017
- 67. Grote, G., & Weichbrodt, J. (2013). Why regulators should stay away from safety culture and stick to rules instead. In C. Bieder & M. Bourrier (Eds.), *Trapping*

safety into rules: How desirable and avoidable is proceduralization of safety? (pp. 225–240). CRC Press. https://doi. org/10.1201/9781315549774-14

- Kwak, J. (2014). Cultural capture and the financial crisis. In D. Carpenter & D. A. Moss (Eds.), Preventing regulatory capture: Special interest influence and how to limit it (pp. 71–98). Cambridge University Press. https://doi.org/10.1017/ CBO9781139565875.008
- 69. Edmondson, A. C. (2012). *Teaming: How organizations learn, innovate, and compete in the knowledge economy.* Jossey-Bass.
- Call, M. L., Campbell, E. M., Dunford, B. B., Boswell, W. R., & Boss, W. (2021). Shining with the stars? Unearthing how group star proportion shapes non-star performance. *Personnel Psychology*, 74(3), 543–572. https://doi.org/10.1111/ peps.12420
- McEwan, D., Ruissen, G. R., Eys, M. A., Zumbo, B. D., & Beauchamp, M. R. (2017). The effectiveness of teamwork training on teamwork behaviors and team performance: A systematic review and meta-analysis of controlled interventions. *PLOS ONE*, *12*(1), Article e0169604. https://doi.org/10.1371/ journal.pone.0169604
- 72. Salas, E., DiazGranados, D., Klein, C., Burke, C. S., Stagl, K. C., Goodwin, G. F., & Halpin, S. M. (2008). Does team training improve team performance? A meta-analysis. *Human Factors*, 50(6), 903–933. https://doi. org/10.1518/001872008X375009
- Weiss, M., Kolbe, M., Grote, G., Spahn, D. R., & Grande, B. (2018). We can do it! Inclusive leader language promotes voice behavior in multi-professional teams. *The Leadership Quarterly, 29*(3), 389–402. https://doi.org/10.1016/j. leaqua.2017.09.002
- 74. Carter, K. M., Mead, B. A., Stewart,
 G. L., Nielsen, J. D., & Solimeo, S. L.
 (2018). Reviewing work team design characteristics across industries:
 Combining meta-analysis and comprehensive synthesis. *Small Group Research*, *50*(1), 138–188. https://doi. org/10.1177/1046496418797431
- 75. Parker, S. K. (2014). Beyond motivation: Job and work design for development, health, ambidexterity, and more. Annual Review of Psychology, 65, 661–691. https://doi.org/10.1146/ annurev-psych-010213-115208
- 76. Grote, G., Kolbe, M., & Waller, M. J. (2018). The dual nature of adaptive coordination in teams: Balancing demands for flexibility and stability. Organizational Psychology Review,

8(2-3), 125-148. https://doi. org/10.1177/2041386618790112

- 77. Humphrey, S. E., Nahrgang, J. D., & Morgeson, F. P. (2007). Integrating motivational, social, and contextual work design features: A metaanalytic summary and theoretical extension of the work design literature. *Journal of Applied Psychology*, 92(5), 1332–1356. https://doi. org/10.1037/0021-9010.92.5.1332
- Kozlowski, S. W. J., Gully, S. M., Salas, E., & Cannon-Bowers, J. A. (1996). Team leadership and development: Theory, principles, and guidelines for training leaders and teams. In M. Beyerlein, D. Johnson, & S. Beyerlein (Eds.), Advances in interdisciplinary studies of work teams: Team leadership (Vol. 3, pp. 251–289). JAI Press.
- 79. Yukl, G. (2013). *Leadership in organizations* (8th ed.). Prentice Hall.
- Pronovost, P., Needham, D., Berenholtz, S., Sinopoli, D., Chu, H., Cosgrove, S., Sexton, B., Hyzy, R., Welsh, R., Roth, G., Bander, & Kepros, J. (2006). An intervention to decrease catheterrelated bloodstream infections in the ICU. New England Journal of Medicine, 355(26), 2725–2732. https://doi. org/10.1056/NEJMoa061115
- Kozlowski, S. W. J., & Chao, G. T. (2018). Unpacking team process dynamics and emergent phenomena: Challenges, conceptual advances, and innovative methods. *American Psychologist*, 73(4), 576–592. https://doi.org/10.1037/ amp0000245
- 82. Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013, October 25). Atypical combinations and scientific impact. *Science*, 342(6157), 468–472. https:// doi.org/10.1126/science.1240474
- Wuchty, S., Jones, B. F., & Uzzi, B. (2007, May 18). The increasing dominance of teams in production of knowledge. *Science*, 316(5827), 1036–1039. https:// doi.org/10.1126/science.1136099
- 84. Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, 46(8), 1416–1436. https://doi.org/10.1016/j. respol.2017.06.006
- Hall, K. L., Vogel, A. L., & Crowston, K. (2019). Comprehensive collaboration plans: Practical considerations spanning across individual collaborators to institutional supports. In K. L. Hall, A. L. Vogel, & R. Croyle (Eds.), *Strategies for team science success* (pp. 587–621). Springer. https://doi. org/10.1007/978-3-030-20992-6_45

editorial policy

Behavioral Science & Policy (BSP) is an international, peerreviewed publication of the Behavioral Science & Policy Association and SAGE Publications. BSP features short, accessible articles describing actionable policy applications of behavioral scientific research that serves the public interest. Articles submitted to BSP undergo a dual-review process: For each article, leading disciplinary scholars review for scientific rigor and experts in relevant policy areas review for practicality and feasibility of implementation. Manuscripts that pass this dualreview are edited to ensure their accessibility to policy makers, scientists, and lay readers. BSP is not limited to a particular point of view or political ideology.

Manuscripts can be submitted in a number of different formats, each of which must clearly explain specific implications for public- and/or private-sector policy and practice.

External review of the manuscript entails evaluation by at least two outside referees—at least one in the policy arena and at least one in the disciplinary field.

Professional editors trained in BSP's style work with authors to enhance the accessibility and appeal of the material for a general audience.

Each of the sections below provides general information for authors about the manuscript submission process. We recommend that you take the time to read each section and review carefully the BSP Editorial Policy before submitting your manuscript to *Behavioral Science & Policy*.

Manuscript Categories

Manuscripts can be submitted in a number of different categories, each of which must clearly demonstrate the empirical basis for the article as well as explain specific implications for (public and/or private-sector) policy and practice:

- Proposals (≤ 2,500 words) specify scientifically grounded policy proposals and provide supporting evidence including concise reports of relevant studies. This category is most appropriate for describing new policy implications of previously published work or a novel policy recommendation that is supported by previously published studies.
- Reports (≤ 3000 words) provide a summary of output and actionable prescriptions that emerge from a workshop, working group, or standing organization in the behavioral policy space. In some cases such papers may consist of summaries of a much larger published report that also includes some novel material such as meta-analysis, actionable implications, process lessons, reference to related work by others, and/or new results not presented in the initial report. These papers are not merely summaries of a published report, but also should provide substantive illustrations of the research or recommendations and insights about the implications of the report content or process for others proposing to do similar work. Submitted papers will undergo BSP review for rigor and accessibility that is expedited to facilitate timely promulgation.

- Findings (≤ 4,000 words) report on results of new studies and/or substantially new analysis of previously reported data sets (including formal meta-analysis) and the policy implications of the research findings. This category is most appropriate for presenting new evidence that supports a particular policy recommendation. The additional length of this format is designed to accommodate a summary of methods, results, and/or analysis of studies (though some finer details may be relegated to supplementary online materials).
- Reviews (≤ 5,000 words) survey and synthesize the key findings and policy implications of research in a specific disciplinary area or on a specific policy topic. This could take the form of describing a general-purpose behavioral tool for policy makers or a set of behaviorally grounded insights for addressing a particular policy challenge.
- Other Published Materials. BSP will sometimes solicit or accept *Essays* (≤ 5,000 words) that present a unique perspective on behavioral policy; *Letters* (≤ 500 words) that provide a forum for responses from readers and contributors, including policy makers and public figures; and *Invitations* (≤ 1,000 words with links to online Supplemental Material), which are requests from policy makers for contributions from the behavioral science community on a particular policy issue. For example, if a particular agency is facing a specific challenge and seeks input from the behavioral science community, we would welcome posting of such solicitations.

Review and Selection of Manuscripts

On submission, the manuscript author is asked to indicate the most relevant disciplinary area and policy area addressed by his/her manuscript. (In the case of some papers, a "general" policy category designation may be appropriate.) The relevant Senior Disciplinary Editor and the Senior Policy Editor provide an initial screening of the manuscripts. After initial screening, an appropriate Associate Policy Editor and Associate Disciplinary Editor serve as the stewards of each manuscript as it moves through the editorial process. The manuscript author will receive an email within approximately two weeks of submission, indicating whether the article has been sent to outside referees for further consideration. External review of the manuscript entails evaluation by at least two outside referees. In most cases, Authors will receive a response from BSP within approximately 60 days of submission. With rare exception, we will submit manuscripts to no more than two rounds of full external review. We generally do not accept re-submissions of material without an explicit invitation from an editor. Professional editors trained in the BSP style will collaborate with the author of any manuscript recommended for publication to enhance the accessibility and appeal of the material to a general audience (i.e., a broad range of behavioral scientists, public- and private-sector policy makers, and educated lay public). We anticipate no more than two rounds of feedback from the professional editors.

Standards for Novelty

BSP seeks to bring new policy recommendations and/or new evidence to the attention of public and private sector policy makers that are supported by rigorous behavioral and/or social science research. Our emphasis is on novelty of the policy application and the strength of the supporting evidence for that recommendation. We encourage submission of work based on new studies, especially field studies (for Findings and Proposals) and novel syntheses of previously published work that have a strong empirical foundation (for Reviews).

BSP will also publish novel treatments of previously published studies that focus on their significant policy implications. For instance, such a paper might involve re-working of the general emphasis, motivation, discussion of implications, and/or a re-analysis of existing data to highlight policy-relevant implications or prior work that have not been detailed elsewhere.

In our checklist for authors we ask for a brief statement that explicitly details how the present work differs from previously published work (or work under review elsewhere). When in doubt, we ask that authors include with their submission copies of related papers. Note that any text, data, or figures excerpted or paraphrased from other previously published material must clearly indicate the original source with quotation and citations as appropriate.

Authorship

Authorship implies substantial participation in research and/ or composition of a manuscript. All authors must agree to the order of author listing and must have read and approved submission of the final manuscript. All authors are responsible for the accuracy and integrity of the work, and the senior author is required to have examined raw data from any studies on which the paper relies that the authors have collected.

Data Publication

BSP requires authors of accepted empirical papers to submit all relevant raw data (and, where relevant, algorithms or code for analyzing those data) and stimulus materials for publication on the journal web site so that other investigators or policymakers can verify and draw on the analysis contained in the work. In some cases, these data may be redacted slightly to protect subject anonymity and/or comply with legal restrictions. In cases where a proprietary data set is owned by a third party, a waiver to this requirement may be granted. Likewise, a waiver may be granted if a dataset is particularly complex, so that it would be impractical to post it in a sufficiently annotated form (e.g. as is sometimes the case for brain imaging data). Other waivers will be considered where appropriate. Inquiries can be directed to the BSP office.

Statement of Data Collection Procedures

BSP strongly encourages submission of empirical work that is based on multiple studies and/or a meta-analysis of several datasets. In order to protect against false positive results, we ask that authors of empirical work fully disclose relevant details concerning their data collection practices (if not in the main text then in the supplemental online materials). In particular, we ask that authors report how they determined their sample size, all data exclusions (if any), all manipulations, and all measures in the studies presented. (A template for these disclosures is included in our checklist for authors, though in some cases may be most appropriate for presentation online as Supplemental Material; for more information, see Simmons, Nelson, & Simonsohn, 2011, *Psychological Science, 22, 1359–1366*).

Copyright and License

Copyright to all published articles is held jointly by the Behavioral Science & Policy Association and SAGE Publications, subject to use outlined in the *Behavioral Science & Policy* publication agreement (a waiver is considered only in cases where one's employer formally and explicitly prohibits work from being copyrighted; inquiries should be directed to the BSPA office). Following publication, the manuscript author may post the accepted version of the article on his/her personal web site, and may circulate the work to colleagues and students for educational and research purposes. We also allow posting in cases where funding agencies explicitly request access to published manuscripts (e.g., NIH requires posting on PubMed Central).

Open Access

BSP posts each accepted article on our website in an open access format at least until that article has been bundled into an issue. At that point, access is granted to journal subscribers and members of the Behavioral Science & Policy Association. Questions regarding institutional constraints on open access should be directed to the editorial office.

Supplemental Material

While the basic elements of study design and analysis should be described in the main text, authors are invited to submit Supplemental Material for online publication that helps elaborate on details of research methodology and analysis of their data, as well as links to related material available online elsewhere. Supplemental material should be included to the extent that it helps readers evaluate the credibility of the contribution, elaborate on the findings presented in the paper, or provide useful guidance to policy makers wishing to act on the policy recommendations advanced in the paper. This material should be presented in as concise a manner as possible.

Embargo

Authors are free to present their work at invited colloquia and scientific meetings, but should not seek media attention for their work in advance of publication, unless the reporters in question agree to comply with BSP's press embargo. Once accepted, the paper will be considered a privileged document and only be released to the press and public when published online. BSP will strive to release work as quickly as possible, and we do not anticipate that this will create undue delays.

Conflict of Interest

Authors must disclose any financial, professional, and personal relationships that might be construed as possible sources of bias.

Use of Human Subjects

All research using human subjects must have Institutional Review Board (IRB) approval, where appropriate.



The Behavioral Science & Policy Association is grateful to the sponsors and partners who generously provide continuing support for our non-profit organization.



SCHOOL of MANAGEMENT

To become a Behavioral Science & Policy Association sponsor, please contact BSPA at bspa@behavioralpolicy.org or 1-919-681-5932.



where behavioral research meets policy + practice

who we are

The Behavioral Science & Policy Association is a global hub of behavioral science resources, curated by leading scholars and policymakers, aimed at facilitating positive change and innovative solutions to a range of societal challenges.



To foster and connect a growing community of interdisciplinary practitioners, providing thoughtful application of rigorous behavioral science research for the public and private sectors, with a simple goal in mind: addressing social change for the benefit of all.

membership

There is a growing movement among social scientists and leaders within the public and private sector, dedicated to grounding important decisions in strong scientific evidence.

BSPA plays a key role in this movement, encouraging decisions to be based on evidence. We need you to join us in this effort to make a lasting impact.

As a BSPA member, you will receive numerous benefits including an oniine subscription to *Behavioral Science & Policy*, early-bird rates for conferences, workshops and briefings, exclusive access to BSPA online webinars and podcasts, waived fees for journal submissions and more.

Be a leader in our drive for change at behavioralpolicy.org/signup

call for submissions

Behavioral Science & Policy is an international, peer-reviewed journal featuring succinct and accessible articles outlining actionable policy applications of behavioral science research that serves the public interest.

BSP journal submissions undergo a dual-review process. Leading scholars from specific disciplinary areas review articles to assess their scientific rigor; while at the same time, experts in designated policy areas evaluate these submissions for relevance and feasibility of implementation.

Manuscripts that pass this dual-review are edited to ensure accessibility to scientists, policymakers, and lay readers. BSPA is not limited to a particular point of view or political ideology. This journal is a publication of the Behavioral Science & Policy Association and SAGE Publications.

We encourage you to submit your manuscript today to *Behavioral Science & Policy*, at behavioralpolicy.org/journal

Behavioral Science & Policy Association P.O. Box 51336 Durham, NC 27717-1336