

a publication of the behavioral science & policy association

bsp

volume 8 issue 1
2022

featured topic

behavioral insights for managing
employee (mis)behaviors

in this issue

a new method for optimizing defaults

behavioralpolicy.org

founding co-editors

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

senior policy editor

Carol L. Graham (Brookings Institution)

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)
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 Michigan)
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 Kruij, Managing Editor (Duke University)
Carsten Erner, Statistical Consultant (FS Card)
A. David Nussbaum, Director of Communications (Chicago)
Ricki Rusting, Editorial Director

consulting editors

Dan Ariely (Duke University)
Shlomo Benartzi (UCLA)
Laura L. Carstensen (Stanford University)
Susan T. Fiske (Princeton University)
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 (University of Texas)
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)
Annmaria 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

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)

Cognitive & Brain Science

Senior Disciplinary Editor
Associate Disciplinary Editors

Henry L. Roediger III (Washington University)
Yadin Dudai (Weizmann Institute & NYU)
Roberta L. Klatzky (Carnegie Mellon University)
Hal Pashler (UC San Diego)
Steven E. Petersen (Washington University)
Jeremy M. Wolfe (Harvard University)

Decision, Marketing, & Management Sciences

Senior Disciplinary Editor
Associate Disciplinary Editors

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 (University of Oregon)
John D. Serman (MIT)
George Wu (University of Chicago)

Organizational Science

Senior Disciplinary Editors
Associate Disciplinary Editors

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)
Ellen E. Kossek (Purdue University)
Elizabeth W. Morrison (NYU)
William Ocasio (Northwestern University)
Sara L. Rynes-Weller (University of Iowa)
Andrew H. Van de Ven (University of Minnesota)

Social Psychology

Senior Disciplinary Editor
Associate Disciplinary Editors

Nicholas Epley (University of Chicago)
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)

Sociology

Senior Disciplinary Editors
Associate Disciplinary Editors

Peter S. Bearman (Columbia University)
Karen S. Cook (Stanford University)
Paula England (NYU)
Peter Hedstrom (Oxford University)
Arne L. Kalleberg (University of North Carolina)
James Moody (Duke University)
Robert J. Sampson (Harvard University)
Bruce Western (Harvard University)

ii

Editors' note

Craig R. Fox & Sim B. Sitkin

Features

1

Finding

How to choose a default

John Beshears, Richard T. Mason,
& Shlomo Benartzi

17

Proposal

Nudge versus sludge in gambling warning labels: How the effectiveness of a consumer protection measure can be undermined

Philip W. S. Newall, Lukasz Walasek,
Elliot A. Ludvig, & Matthew J. Rockloff

25

Finding

Encouraging COVID-19 vaccination through behaviorally informed reminders: Results from a national randomized field experiment in Israel

Adi Berliner Senderey, Reut Ohana,
Shay Perchik, Ido Erev, & Ran D. Balicer

35

Finding

Ventilator allocations: The effect of mere identifiability

Ilana Ritov & Stephen M. Garcia

47

Field Review

Encouraging employees' active feedback & participation when rolling out major changes

Elizabeth A. Hood & Jean M. Bartunek

61

Field Review

"Braking" bad: How managers can respond to employee misbehavior

Erin L. Frey, Gabrielle S. Adams,
Evan A. Bruno, & James R. Detert

78

Editorial policy

Welcome back to the pages of *Behavioral Science & Policy*, which promotes the thoughtful application of rigorous behavioral science to policy and practice in ways that serve the public interest. The current issue features contributions that cluster around three topics: choice architecture, public health, and organizational policies.

Our first two articles explore ways to enhance the effectiveness of different forms of choice architecture.

As applications of nudging spread around the world, choice architects have begun to examine the heterogeneous effects of interventions on targeted individuals. For instance, a hospital administrator may wish to decrease the prescription of potentially addictive opioids by programming an electronic medical record system to set the default order to 10 pills. Although this default is likely to reduce the number of pills prescribed by physicians who would normally prescribe more liberally, it may also have the unintended effect of increasing the number of pills prescribed by a physician who would normally prescribe more conservatively. More generally, when policymakers are confronted with ordered sets of options (such as the proportions of income employees set aside for retirement), they face a dilemma of where to set a one-size-fits-all default to maximize overall impact. John Beshears, Richard T. Mason, and Shlomo Benartzi address this question by showing how choice architects can model the impact of defaults on the distribution of latent preferences in a population (that is, choices made the absence of a preset default) and calibrate these models to determine what the optimal default might be.

A second case in which well-meaning nudge policies can have potentially unintended consequences involves warning labels. For instance, regulators in the United Kingdom require operators of online gambling games (for example, roulette and slot machines) to post information for consumers about the average loss from each gamble. However, operators are given considerable discretion in how to

interpret and implement these mandates. Philip W. S. Newall, Lukasz Walasek, Elliot A. Ludvig, and Matthew J. Rockloff report that operators tend to present such warning labels in ways that render them systematically less useful to consumers than regulators intend them to be. For instance, average loss statistics are almost always reported as the percentage of staked money paid out in winnings, a “return-to-player” framing that has been shown to be confusing to consumers, rather than the more understandable “house-edge” framing, which indicates what the house keeps. Moreover, the authors note that these warning labels tend to be difficult for consumers to access, buried in wordy information statements, and presented in small and nondistinctive fonts. The authors conclude with recommendations for regulators to help enhance the impact of warning labels.

As the COVID-19 pandemic transitions to an endemic phase, it is important to take stock of lessons from behavioral science research for managing future health crises. Two articles in this issue are novel entries for this literature (for more on this topic, see the second issue of Volume 6 of *Behavioral Science & Policy*, from fall 2020).

One stubborn challenge for managing pandemics and other disease outbreaks is persuading people to get vaccinated against them. Adi Berliner Senderey, Reut Ohana, Shay Perchik, Ido Erev, and Ran D. Balicer present the results of a large-scale field study in which several hundred thousand Israelis received a text message reminding them to get vaccinated. The text consisted of either (a) a simple (control) message, (b) a message that invited them to join the large number of vaccinated citizens (designed to leverage the social norm as an influence tactic), or (c) a message citing research on the health benefits of the vaccine. Both experimental messages substantially increased vaccination rates over the subsequent eight days. Interestingly, the health benefits message was slightly (but reliably) more effective than the social norm message in this situation. Although these results may or may not generalize to different

populations or for different vaccines, this study could provide a promising starting point for future vaccination efforts.

Ilana Ritov and Stephen M. Garcia take on a challenging topic that troubled health care administrators early in the pandemic and during the omicron surge: How should the limited number of ventilators be allocated to a growing number of very sick patients? There is broad agreement that younger patients generally ought to be prioritized over older patients to save more future years of life. Of course, this is easier said in the abstract than done in practice when patients are seen as individuals rather than statistics. Indeed, the authors show in three studies that participants were less likely to say they would reassign a ventilator from an older patient to a younger patient when the patients had been identified by name than when they had not. Apparently, modest identifiability is all that is needed to induce an emotional response. Anonymity may be the key to more rational decision-making in such cases.

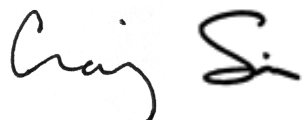
Our final two articles in this issue speak to behaviorally informed organizational policies.

Elizabeth A. Hood and Jean M. Bartunek explore how employees respond to organizational change. Leaders naturally tend to focus on how to promote acceptance and minimize rejection of change initiatives among employees. The authors argue, counterintuitively, that leaders ought to instead focus on how actively employees engage with these initiatives. When employees passively accept a new program, they do not raise questions, provide feedback,

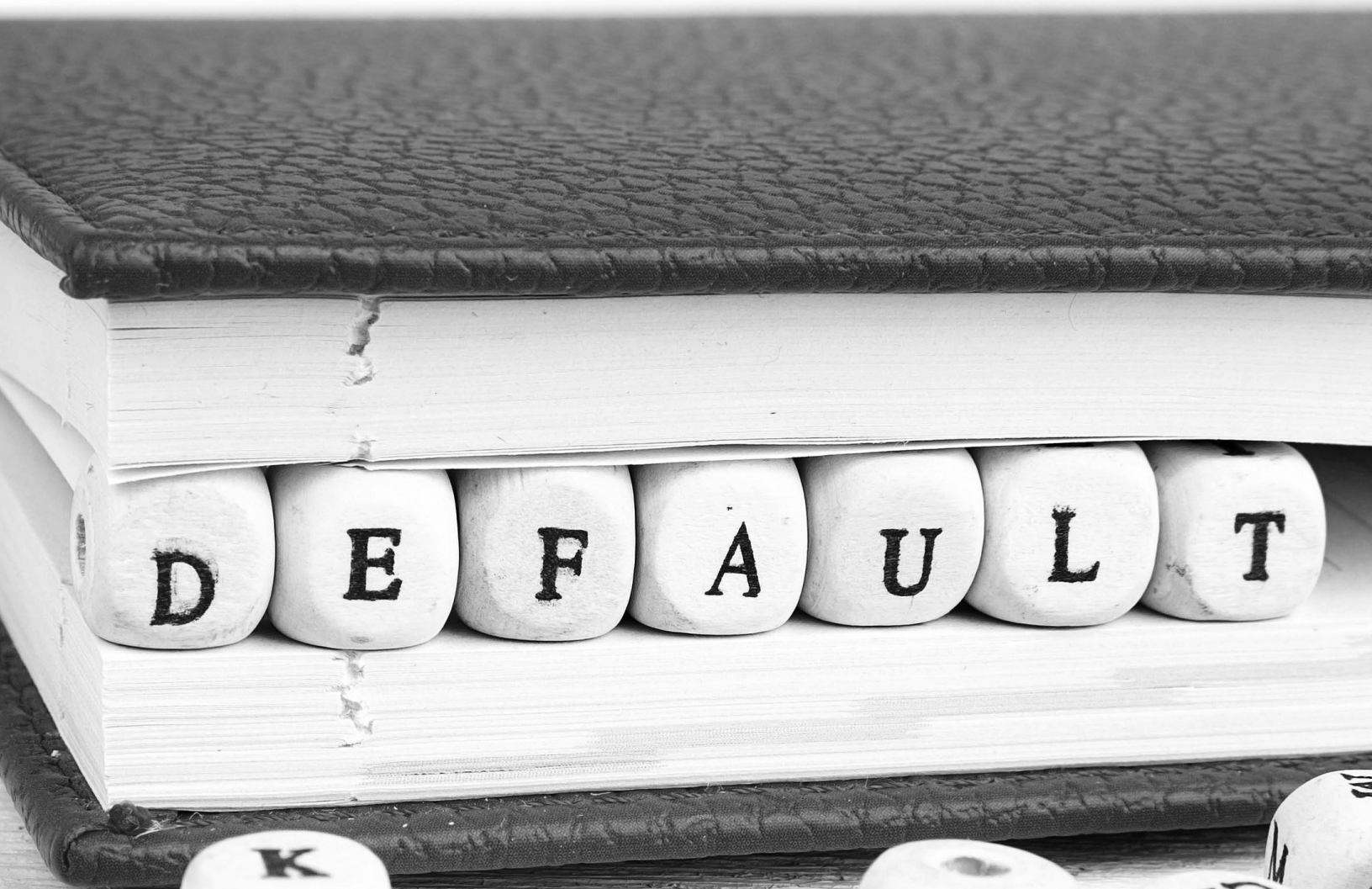
and troubleshoot. Thus, the organization misses opportunities to enhance the new policy and its implementation. Likewise, when employees passively disengage from a new initiative, they may make mistakes, drag their feet, and become more cynical about their jobs. Again, the organization misses opportunities to identify ways to improve employee performance and morale. The authors illustrate these scenarios with concrete case studies and derive lessons for how managers can more effectively lead organizational change.

The last article in this issue, by Erin L. Frey, Gabrielle S. Adams, Evan A. Bruno, and James R. Detert, addresses how managers can most effectively respond to employee misbehavior that neither is as benign as unintentional mistakes nor rises to the level of criminal misconduct. Although organizations often have policies that demand some sort of action by managers, such as documentation and warnings, they typically offer little guidance on how to best implement these actions. In this article, the authors mine organizational behavior research to derive guidelines for selecting and communicating consequences for misbehaviors in ways that balance the goals of serving justice, appearing fair and compassionate, and deterring future offenses.

Thank you for your interest in *Behavioral Science & Policy*. As always, we welcome feedback from the community and look forward to continuing to share with you the latest insights from behavioral science research for policy and practice.



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



How to choose a default

John Beshears, Richard T. Mason, & Shlomo Benartzi

abstract

We have developed a model for setting a default when a population is choosing among ordered choices—that is, ones listed in ascending or descending order. A company, for instance, might want to set a default contribution rate that will increase employees' average contributions to a retirement savings plan. A key input of the model is the distribution of *latent* options—the percentages of a population that select each available choice in the absence of a preset default. The model treats the default as an attraction point that causes some people to shift from their latent preference toward the default. It specifies the strength of each possible default's pull on each latent option and thereby points policymakers to the default most likely to achieve a desired aim. We tested our model using data from field experiments relating to retirement savings. In addition to presenting the results, which support the model's validity, we discuss how the model relates to prior empirical evidence on defaults.

Beshears, J., Mason, R. T., & Benartzi, S. (2022). How to choose a default. *Behavioral Science & Policy*, 8(1), 1–15.

When asked to identify the greatest contribution of behavioral economics to policy, Richard H. Thaler, winner of the Nobel Memorial Prize in Economic Sciences for his work on behavioral economics, pointed to improvements in retirement savings policies. One of the improvements involved changing the default from nonparticipation to participation in defined contribution plans, in which employees set aside a given amount for the plan from each paycheck.¹ Behavioral science research has documented that participation in such retirement plans increases dramatically when companies switch from an opt-in plan (in which no contribution is made unless an employee actively selects a contribution rate) to an opt-out plan (in which a preset, or *default*, contribution rate is used unless an employee actively changes the rate).^{2–8} In a 2016 survey, 60% of 401(k) plans indicated that they use opt-out policies,⁹ and such policies have been implemented at the national level in the United Kingdom, New Zealand, and Turkey.

The use of defaults is not limited to retirement plans. Changing the default changes the actions people take in domains such as organ donation,^{10–12} insurance,¹³ online marketing,¹⁴ consumer product choice,¹⁵ energy use,¹⁶ tipping,¹⁷ medication prescriptions,^{18,19} and charitable donations.^{20,21} A recent meta-analysis, combining data from multiple studies, documents the effects of defaults across settings.²²

In this article, we offer guidance to policymakers who must choose a default from among many ordered options (see note A)—that is, ones listed in rank order, such as the percentage of salary contributed to a retirement plan. We present a model for determining whether and by how much any given default will cause individuals to deviate from their *latent* choice (the one that would be selected in the absence of a default) and for deciding which default to set for the greatest good of a population.

To aid readers in understanding the model, we speak mostly in terms of the concrete example of how it would be used for increasing contributions to defined contribution retirement plans. After describing the model, we report on field

studies that illustrate how the model applies in a particular setting and provide empirical support for the model's predictions.

Model Basics

In applying the model to retirement plans, we take the perspective of a company policymaker who wishes to shift employee contribution rates upward (see note B). The policymaker may believe, for example, that a group of individuals is saving too little.⁸ The insights that we develop in the context of retirement plans could help address challenges in other domains, such as encouraging a group of workers to lower their chosen thermostat settings in the winter¹⁶ or inducing a group of physicians who are overprescribing brand-name medications to switch to prescribing generic equivalents.^{18,19}

In a nutshell, we assume that a default is an attraction point. Any given default might cause some percentage of people to shift from their latent preference toward the default. When applying the model to a defined contribution retirement plan, we first collect data describing the distribution of latent contribution rates: For a subset of people in the population of interest, we observe the rate choices made in the absence of a default, and we identify the percentages who chose each latent value. Then we use the model to calculate the percentages of people who will shift from their latent value and how far they will shift in response to each of several defaults, with the resulting distributions of choices varying depending on different assumptions about the strength of the default's pull on nearby and distant latent values. The strength of the default is modeled by a combination of factors, or *parameters*, each of which can take any value in a range of values.

We then identify the combination of parameters that generates the model's most accurate predictions for how the distribution of contribution rates changes when a default is introduced. For a given combination of parameters, we compare the model-based predictions with actual choices made by a group of employees who demographically resemble the group that saw no default but were presented with a default. By finding

Core Findings

What is the issue?

Changing the default option—the option that is implemented on behalf of individuals who do not actively select an option for themselves—has been a major success for behavioral scientists seeking to influence policy-relevant outcomes. Choosing defaults, however, is difficult because the effects of defaults are nuanced. We provide guidance for choosing a default among ordered options when the goal is to increase the mean outcome in a population.

How can you act?

Selected recommendations include:

- 1) Examining experimental or nonexperimental data, as available, to determine the strength of defaults
- 2) When defaults are weak, placing a default just above a cluster of options that many individuals would select if forced to make a choice
- 3) When defaults are strong, placing a default well above such a cluster of options

Who should take the lead?

Policymakers and organizational leaders

the model-based predictions that most closely match the real-life choices that people make in response to several defaults, we identify the parameter values that best describe the population's responsiveness to defaults. Using those parameter values, we calculate the default that would be most effective at shifting the population's mean selection in a desired direction. (The model can also be useful when a policymaker's ability to collect data is limited. We address this situation in the article's last section.)

In principle, a policymaker seeking to increase the mean contribution rate of a group of employees should select a default that pulls many contribution rates up while pulling few down. Our modeling indicates that to achieve this aim, a policymaker should choose a default that is above a cluster of popular latent contribution rates. When the default has a weak influence, the default that maximizes the mean contribution rate is likely to be one that is only slightly above that cluster. When the default has a strong influence, policymakers will maximize the mean contribution rate by setting a default that is significantly above the cluster. To more fully explain how the model guides policymakers to a desirable default, we provide concrete details in the next section and in the Supplemental Material. We should note that we consider our model to be empirical rather than explanatory, because it does not specify the economic and psychological mechanisms driving people's responses to defaults; it merely describes how different defaults shape the distribution of the selections made by a population. As we explain later, however, its assumptions are consistent with the findings of past behavioral science research into defaults.

The Model in Detail

Our model has three ingredients: (a) the distribution of latent contribution rates, (b) the value of a default, and (c) a set of formulas that uses three parameters to predict whether and to what extent the default will affect an individual who has a particular latent contribution rate.

In applying the model to retirement savings plans, we obtained the distribution of latent

"In a nutshell, we assume that a default is an attraction point"

choices from data we collected in one of three empirical studies we conducted (these are described more fully later in the article). In this study—which we call Study 1 despite its not having been done first—one group of employees visited a website to select their contribution rate to a company's retirement program and saw no default; they used a keyboard to enter a number into a blank space. After entering a number, they could either retain that initial number or select a different number, whether by interacting with the website during the same session or by returning to the website at a later time. We defined an individual's latent contribution rate (L) as the individual's selected rate at the end of this process. Other groups of employees in Study 1, as well as employees in our two other empirical studies (Studies 2 and 3), saw a prepopulated number where employees in the first group saw a blank space. Otherwise, their experience was identical to that of employees in the first group. The prepopulated number is the default value (D), and in our studies, it took values in the range of 6% to 11%.

The third ingredient—the heart of the model—uses two formulas to govern individuals' responses to defaults, with the three parameters mentioned above determining how strongly defaults influence choices.

The first formula predicts the likelihood that an individual with a given latent choice will end up changing their choice in response to a given default. It incorporates a parameter termed the *radius* (R) and an *adjustment factor* (F).

The R value indicates how far a default's influence extends. If the default were a point on a horizontal line indicating the potential contribution rates in ascending order, the radius would describe the distance to the right and to the left within which the default has an effect on latent rates. For instance, an R of 4 around a default of 7% indicates that the default has an effect on

individuals whose latent values are between 3% and 11%.

The formula reflects the presumption, based on past research into defaults, that a default's effect declines as the latent rate goes in either direction from the default. When a person's latent contribution rate is close to the default, the default is likely to cause the person to shift from the latent rate in the direction of the default. When a person's latent contribution rate is far from the default, the default is less likely to influence the outcome. In general, when the default is above an individual's latent contribution rate, the default is likely to pull the individual's contribution rate higher. Conversely, when the default is below the latent contribution rate, the default may pull down the final contribution rate selection.

60%

401(k) plans that indicated use of opt-out policies in 2016

Heart of the model

2 formulas that represent individuals' responses to defaults

psychological anchoring

Tendency for a person asked to choose a numerical value to start with some reference point and then only slightly adjust away from it

In the model, an individual whose latent contribution rate is equal to the default is deemed to be 100% likely to be influenced by the default. An individual whose latent contribution rate is a distance of exactly R away from the default is deemed to have zero probability of being influenced by the default. The probability that the default has an effect declines linearly between those two points, except when the latent contribution rate is especially attractive on its own, an issue that we discuss next. (See note C for a caveat.)

The parameter F is an adjustment factor for the counterpull exerted by latent values that are especially attractive on their own. In retirement savings plans, our own research and other data sets have shown that people like to choose contribution rates that are multiples of five, and they tend to resist shifting to other values; that is, in common with defaults, multiples of five are attraction points. We adjust for this counterpull by decreasing the calculated pull of the default by the amount of resistance generated by latent multiples of five. In the model, we address the counterpull with a formula, or *function*, called $A(L)$. $A(L)$ is set to zero if the latent contribution rate L is not a multiple of five, because these latent rates are assumed to exert no resistance. For multiples of five, $A(L)$ is set to F , with F taking a value in the range from 0 to 1 ($0 \leq F \leq 1$).

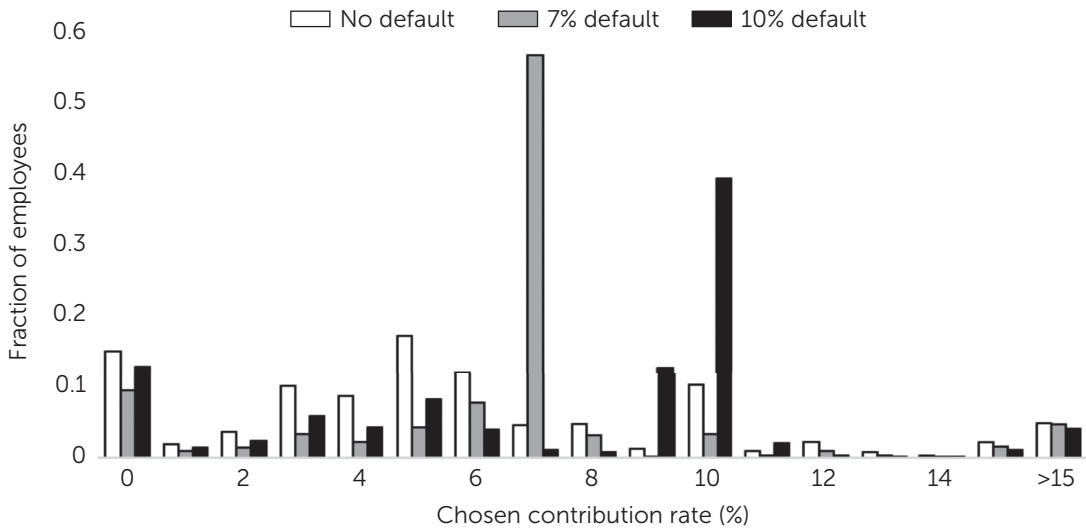
In mathematical notation, we say that when $|D - L| \leq R$ (that is, when the difference between the default value D and the latent contribution rate L is less than or equal to the parameter R), the probability that the individual is influenced by the default is $[1 - A(L)] \times (1 - |D - L|/R)$. When $|D - L| > R$, the probability that the individual is influenced by the default is zero.

If an individual is influenced by the default, the individual ends up with a contribution rate calculated by the second formula at the heart of our model: $[(1 - W)L + WD]$. This formula yields a weighted average of the latent contribution rate and the default. The weighting factor (W) takes a value between 0 and 1 ($0 \leq W \leq 1$). For example, a W of 0.7 (giving fairly strong weight to the default) paired with a D of 10 and an L of 0 would result in a final contribution rate of 7, with the individual moving seven tenths of the way from their latent value toward the default. The online interface in our three studies encouraged individuals to choose contribution rates as whole number percentages, so we round the contribution rate given by the formula to the nearest whole number.

To illustrate how the model works, we have constructed two examples. In the first, depicted in Figure 1, the model uses the parameters $R = 12$, $F = 0.3$, and $W = 0.9$, a combination indicative of the default having a strong, far-reaching influence and multiples of five exerting a modest resistance. The white bars reflect the distribution of actual latent contribution rates revealed by participants in Study 1. The gray and black bars show the predicted distribution of contribution rates when the default was modeled at 7% or 10%, respectively. The predictions indicate that when a default exerts a strong effect over a wide range of latent values, a wide swath of the distribution will be drawn toward the default. Notably, regardless of people's latent rates, a high percentage of those who were modeled to have been presented with a default of 7% switched their choices, and the percentage who chose 7% rose from less than 5% to more than 50%.

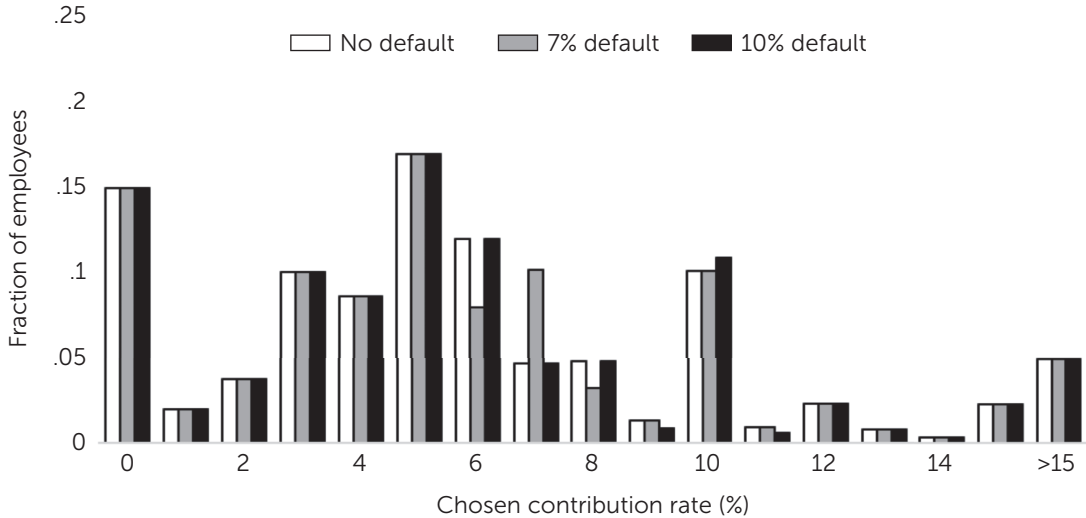
The second example uses the same defaults and distribution of latent contribution rates as in the first example (see Figure 2). The defaults'

Figure 1. Illustrative model prediction of contribution rates when a default exerts a strong effect on rates (parameter values: $R = 12$, $F = 0.3$, $W = 0.9$)



Note. The white bars show the distribution of contribution rates in the no default condition (that is, the latent rates); those rates are drawn from the empirical data in Study 1. The gray and black bars show the distribution of contribution rates predicted by the model for a 7% default and for a 10% default, respectively. The results indicate that when a default exerts a strong effect, people having latent preferences both near to and far from the default will be drawn toward the default—as is evidenced by the declines in the fraction of employees choosing many of the rates. See the main text for definitions of R , F , and W .

Figure 2. Illustrative model prediction of contribution rates when a default exerts a weak effect on rates (parameter values: $R = 1.5$, $F = 0.3$, $W = 0.7$)



Note. As in Figure 1, the white bars show the distribution of contribution rates in the no default condition (that is, the latent rates), and the gray and black bars show the distribution of contribution rates predicted by the model for a 7% default and for a 10% default, respectively. The results indicate that even when the 7% default has a weak effect, it nonetheless exerts a draw on individuals whose latent contribution rates are 6% or 8% (as indicated by the decline in the fraction of employees predicted to choose those rates). The 10% default has less of an effect on the distribution. The parameter values used here generate predictions that most closely match (that is, are the best fit for) the empirical findings from Studies 1, 2, and 3—a correspondence implying that the parameter combination is the best for predicting the responses to defaults in a real-life population resembling that in our studies.

effects are assumed to be weaker, however, as is reflected in the parameters $R = 1.5$, $F = 0.3$, and $W = 0.7$. The output implies that despite the weak effect of the 7% default (gray bars), this default still exerts a draw on individuals whose latent contribution rates are 6% or 8%. The 10% default (black bars) has less of an impact on the distribution. It exerts a pull on individuals whose latent contribution rates are 9% or 11%, but such an influence is less meaningful because few individuals have those latent contribution rates.

This last combination of parameters yielded the best fit with our experimental data; that is, it most closely replicated the outcomes we found when the real-life employees we studied were presented with a default of 6%, 7%, 8%, 9%, 10%, or 11%. Later in the article, we address the implications of this finding for setting defaults, but first we describe the empirical studies we conducted and the ways in which they support the validity of our model and confirm past research on defaults.

The Experiments

Experimental Design

We conducted three experiments, all of which were completed before we conducted our modeling. Even though the experiment we call Study 1 was not run first, we treat it as our primary study because it was the only one that enabled us to observe the distribution of latent

contribution rates and thus to study the effects of various defaults on that distribution.^{20,21,23–25} We describe Study 1 in this subsection and address Studies 2 and 3, which were similar, in a later subsection. For details, see the Supplemental Material.

We worked with the segment of the company Voya Financial that helps employers manage retirement savings plans. For a subset of employers, employees who became eligible for the retirement plan were invited to visit a Voya-administered website, Voya Enroll, to begin contributing. Figures S1–S8 in the Supplemental Material show screenshots of what employees saw during the registration process.

In Study 1, we assigned employees randomly to one of three groups when they reached the webpage at which they selected their contribution rates. The study had three conditions: 7% default, 10% default, and no default.

As shown in Table 1, approximately half of the individuals in Study 1 were men. The mean age of participants was 38 years, and their mean annual salary was approximately \$70,000. These characteristics did not show statistically significant differences across the three conditions.

In the 7% default and 10% default conditions of Study 1, the space for indicating the desired contribution rate was prepopulated with the

Table 1. Participant gender, age, & salary in Study 1, by randomly assigned condition

Characteristic	Experimental condition			<i>p</i> value from χ^2 or <i>F</i> test for null hypothesis that conditions are equal
	No default	7% default	10% default	
% men	53	52	52	.66
Age in years				.69
Mean	38	38	38	
Standard deviation	12	12	12	
Salary				.16
Mean	\$69,000	\$71,000	\$71,000	
Standard deviation	\$51,000	\$52,000	\$54,000	
Number of participants	3,991	4,024	4,048	

Note. The *p* values indicate that the three participant groups do not differ significantly. *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.

default of interest (see Figure S4 in the Supplemental Material). In the no default condition, the space for indicating the desired contribution rate was empty when the webpage loaded, and a blinking cursor prompted the employee to enter a number. (See Figure S5 in the Supplemental Material.) As soon as a number was entered, the webpage transformed to appear as if the entered number had been the prepopulated contribution rate (as in Figure S4 in the Supplemental Material). In all three conditions in Study 1, employees could increase or decrease their chosen contribution rate away from the initial rate by clicking on the + or – keys.

As specified when we preregistered Study 1 (see Figure S9 in the Supplemental Material), our primary outcome variable is the contribution rate in effect 60 days after the initial Voya Enroll visit, adjusted to reduce the potentially misleading influence of outliers by setting values below the 1st percentile equal to the 1st percentile and values above the 99th percentile equal to the 99th percentile. Preregistration is done for transparency, that is, to minimize the likelihood that researchers will cherry-pick data and thus publish misleading results. The choice of a 60-day window balances two factors. On the one hand, a longer time window would increase the likelihood that factors unrelated to the default, such as salary increases or financial emergencies, could influence the final contribution rates. On the other hand, a shorter time window might miss changes that employees make after having some time to ponder their choice more fully. Some employees choose not to enroll in the plan when they first visit Voya Enroll but return within a few weeks and select a positive contribution rate.

Results

As we pledged in our preregistered analysis plan for Study 1, we calculated, using the analytic method known as *ordinary least squares regression*, the effect of the 7% or 10% default on the mean contribution rate. Relative to having no default, the 7% default decreased the mean contribution rate by 0.02 percentage points when we did not control for gender, age, and salary and by 0.04 percentage points when we accounted for those factors; the 10% default

“defaults can trigger shifts from latent values”

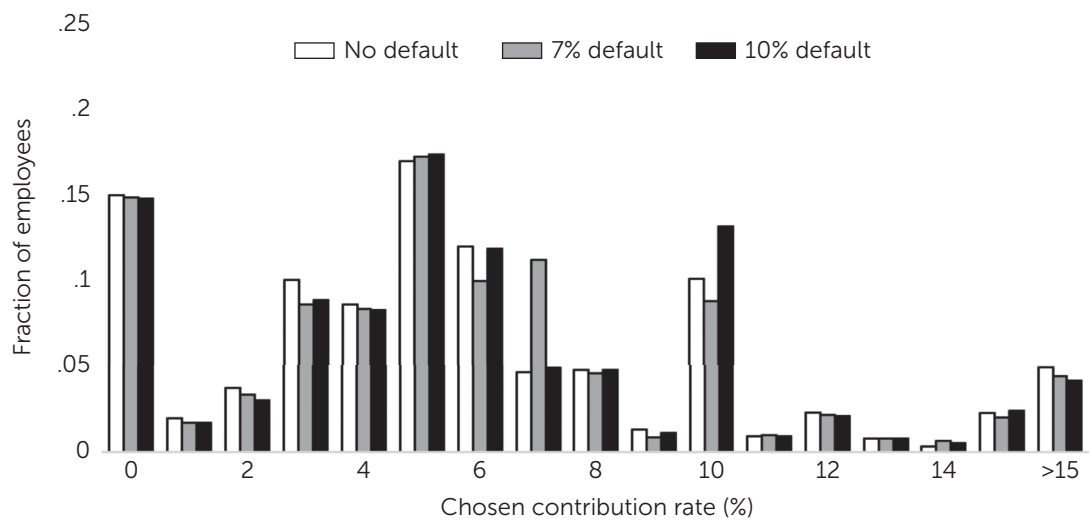
increased the mean contribution rate by 0.08 percentage points when we omitted controls and by 0.06 percentage points when we included controls. None of these estimates were statistically significant, and all of them were small in magnitude. When we used the same analytic approach to investigate whether the 7% default and 10% default increased the likelihood that an individual would choose a contribution rate greater than zero, we similarly found that the effects were not statistically significant and were small in magnitude. We had hypothesized that the 7% default and 10% default would increase a population's mean contribution rate relative to having no default, so we were surprised by these results.

Such findings could have implied that setting defaults did not influence contribution decisions, but further analyses, which were not preregistered, indicated that defaults did, indeed, affect contribution decisions, even though they did not affect mean contribution rates. The results support the idea that defaults can trigger shifts from latent values among people who are signing up for retirement plans even when the average rate for the population does not change in a desired direction.

The data from Study 1 also showed that study participants were attracted to contribution rates that were multiples of five, as previous work has found.⁴ This attraction is evident in Figure 3, which shows the distribution of the final contribution rates in the three experimental conditions. It is because of this finding that our model assumes that individuals whose latent contribution rates are multiples of five are less likely than others to be influenced by defaults. (See note D.)

Additional analyses revealed specific influences of defaults, including that the defaults in the study increased the fraction of individuals who ended up with contribution rates equal to the default. To identify this pattern, we compared

Figure 3. Distributions of contribution rates by randomly assigned condition in Study 1



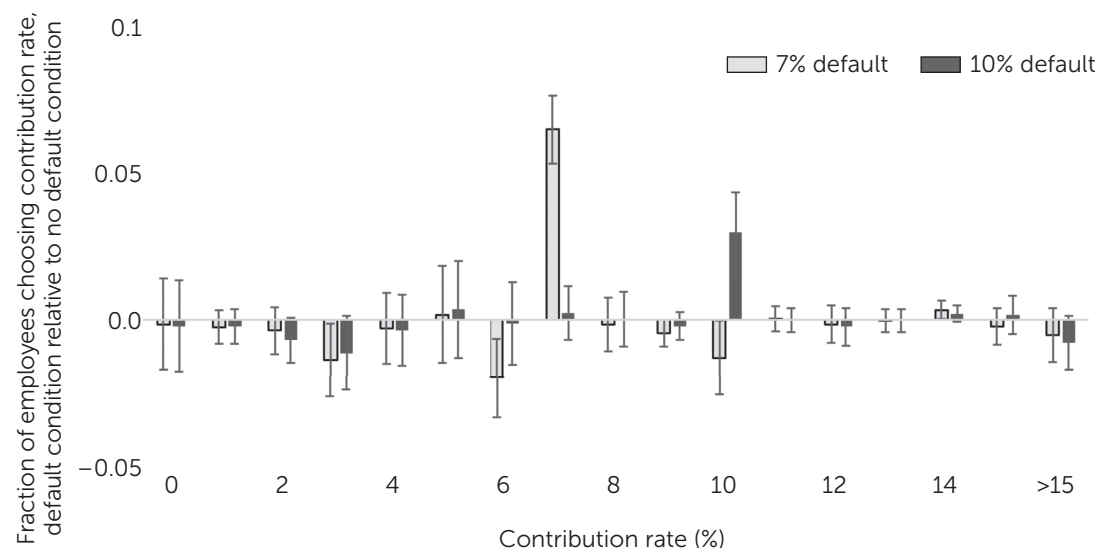
Note. The white, gray, and black bars show the distribution of contribution rates in Study 1 in the no default condition, the 7% default condition, and the 10% default condition, respectively. The data reflect a relatively weak influence of the defaults but show some shifting of latent values toward the defaults—as is indicated by increases in the fraction of employees selecting the defaults and declines in the fractions choosing several other rates.

the fraction of employees who chose a given contribution rate (termed $C\%$, with C being an integer) in the 7% default and 10% default conditions with the fraction of employees who chose that rate in the no default condition. Using ordinary least squares regression again, we calculated the differences separately for each integer contribution rate from 0% to 15%. Because very few people select higher contribution rates, we treated all contribution rates equal to 16% or higher as belonging to a single category. In Figure 4, the contribution rate varies along the horizontal axis. The light and dark vertical bars indicate, respectively, the effect of the 7% default or the 10% default on the likelihood of a given contribution rate being chosen, relative to the likelihood when no default was presented. This effect is measured in terms of the size of the difference in the percentage of employees who chose the given contribution rate. The I-shaped lines, commonly known as *whiskers*, give 95% confidence intervals; the findings are statistically significant when the whiskers do not pass through the horizontal zero line. (See note E for a definition of 95% confidence intervals.)

The data indicate that relative to having no default, the 7% default caused a statistically significant increase in the fraction of individuals with a 7% contribution rate, and the 10% default caused a statistically significant increase in the fraction of individuals with a 10% contribution rate. This finding, too, supports our model, which predicts that some individuals with latent contribution rates close to a default will end up choosing the default. This finding is also consistent with prior literature documenting that defaults are chosen frequently.^{2–6,17,20,21}

The analyses also revealed—again consistent with our modeling and with previous findings^{5,6,20,21}—that individuals sometimes ended up choosing the default either when their latent contribution rate was below the default or when their latent contribution rate was above the default. The 7% default decreased the fraction of employees with a contribution rate less than or equal to 6% and decreased the fraction of employees with a contribution rate greater than or equal to 8%. The 10% default decreased the fraction of employees with a contribution rate less than or equal to 9% and decreased the fraction of employees with a contribution rate greater than or equal to 11%, although the last

Figure 4. The effect of the 7% default & the 10% default on the likelihood of the contribution rate being equal to a given value in Study 1



Note. The plot here compares the fraction of employees who chose a given contribution rate in the 7% default or 10% default condition with the fraction of employees who chose that rate in the no default condition. The results indicate that the defaults increased the fraction of individuals who ended up with contribution rates equal to the default. The data support our model, which predicts that some individuals with latent contribution rates close to a default will end up choosing the default. The whiskers show 95% confidence intervals; findings are statistically significant when the whiskers do not pass through the horizontal zero line.

finding was not statistically significant, perhaps because the fraction of employees with a latent contribution rate greater than or equal to 11% is so low that there is little room to decrease it further.

Finally, in line with our modeling, we found some evidence consistent with past research that indicated a default is more likely to influence an individual whose latent contribution rate is close to the default^{5,6,17,20,21}—although the statistical power of the tests we did is low. (See note F. Also see the Additional Statistical Tests section of the Supplemental Material for more detail about the analyses relating to this finding.)

Calibrating the Model

To see if the model could be useful for indicating which default would be best for raising the mean contribution rate of a given population, we used the following process. To *calibrate the model*, we examined the effects of all possible combinations of R , F , and W on the distributions of contribution rates in the presence of default

rates and looked for the distributions that best matched the real-world distributions found in our empirical studies. The combination of parameter values that led to the best-fit distributions could be presumed to predict the behavior of other populations whose demographic characteristics were similar to those of the study participants when those populations encountered retirement sign-up programs similar to those our participants encountered.

To find the best fit, we combined data from Study 1 with data from Studies 2 and 3. Studies 2 and 3 were conducted prior to Study 1 and were not preregistered. They had the same design as Study 1 except that they lacked a no default condition and had conditions with integer default contribution rates of 6% through 11%, rather than solely 7% and 10%. See the Supplemental Material for more details.

Given the distribution of latent contribution rates from the no default condition of Study 1, we input all possible combinations of the parameters into the model. R ranged from 0.5 to 15.0 in increments of 0.5; F ranged from 0 through 1.0

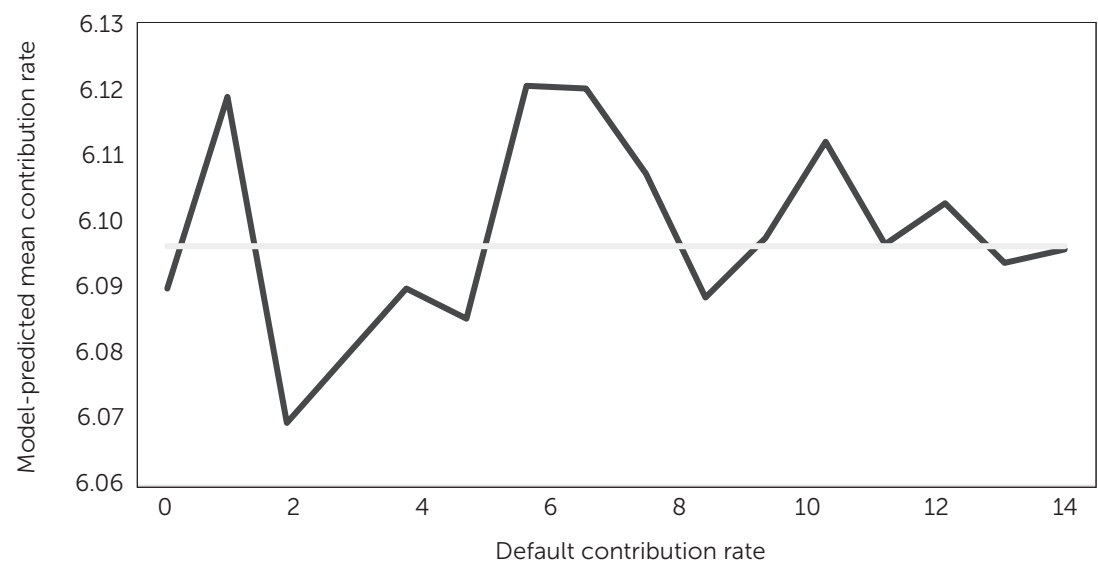
in increments of 0.1; and W ranged from 0 to 1.0 in increments of 0.1. Using each combination, the model predicted the distribution of contribution rates when employees were presented with a 6%, 7%, 8%, 9%, 10%, or 11% default. We then compared the model's predictions with the observed distributions of contribution rates of the participants, and we calculated how closely the modeled distributions matched those found in the empirical studies. See the Supplemental Material for more details.

As we mentioned earlier, for the default values examined in our studies (6%, 7%, 8%, 9%, 10%, and 11%), the model best fits the data when R takes a value of 1.5, F takes a value of 0.3, and W takes any value in the interval $0.5 < W \leq 1.0$ (with $R = 1.5$, the model makes the same predictions for all of these values of W).

Of course, the model's predictions using the best-fitting parameter values do not capture every feature of the real-world data. For example, the model with these parameter values predicts that the default does not affect

individuals whose latent contribution rates are two percentage points or more away from the default. However, in the data from Study 1, the 7% default condition leads to a statistically significant 2.2 percentage point decrease in the fraction of individuals who choose contribution rates of 5% or less and a statistically significant 2.4 percentage point decrease in the fraction of individuals who chose contribution rates of 9% or more, relative to the no default condition (because those individuals moved toward the default). Similarly, the 10% default condition leads to a statistically significant 2.2 percentage point decrease in the fraction of individuals who chose contribution rates of 8% or less, relative to the no default condition. (The 10% default condition did not have a statistically significant effect on the fraction of individuals who chose contribution rates of 12% or more, relative to the no default condition.) Overall, though, the best-fitting parameter values for the model include a low value of R , indicating that whatever the default rate is, it tends to attract individuals whose latent contribution rates are close to the default.

Figure 5. Mean contribution rate predicted by the model with best-fit parameter values ($R = 1.5$, $F = 0.3$, $W = 0.7$) as the default varies



Note. Using the model parameters that produced contribution-rate distributions most like those in the empirical studies ($R = 1.5$, $F = 0.3$, $W = 0.7$), we determined that setting a default of 6% or 7% would result in the highest mean contribution rate in a population that resembled the one in our empirical studies. The horizontal line in the middle of the figure shows the mean contribution rate in the no default (that is, latent) condition. To arrive at the means shown, we calculated the model's predictions for the distribution of contribution rates in response to each possible integer default and then computed the mean of that distribution. One benefit of the model is that it makes predictions about contribution-rate distributions for defaults that we did not test in our experiments (defaults less than 6% or greater than 11%).

Figure 5 shows the model's predictions, given the best-fit parameter values, for the mean contribution rate as the default varies. The model-predicted mean reaches a peak at a default of 6%, and the mean for a default of 7% is nearly identical. In other words, to maximize the mean contribution rate of a population that resembles the one in the studies, the model indicates that policymakers would set a 6% or 7% default. (See note G.) The mean contribution rate for a population is calculated as follows: For each possible contribution rate, we multiply the contribution rate by the fraction of the population predicted by the model to choose that contribution rate, then we calculate the sum across contribution rates.

Comparisons With Other Models

As we mentioned at the start of this article, we constructed our model without specifying the mechanisms driving individuals' responses to defaults. Nonetheless, the model can be compared with ones that articulate mechanisms for the effects of defaults.

Our model implies that individuals whose latent contribution rates are closer to the default are more likely to be affected by the default, a feature consistent with models assuming that people incur a cost—in the form of inconvenience—if they opt out of a default.^{21,23,26} In these models, individuals' latent contribution rates are assumed to be their most preferred contribution rates; therefore, individuals whose latent contribution rates are farthest from the default have the strongest incentive to bear the inconvenience of opting out and switching from the default to their most preferred contribution rates. Individuals whose latent contribution rates are close to the default have a weaker incentive to go to the trouble of opting out and are more likely to remain at the default.

Yet our model differs from those that focus on the costs of opting out in that, like models based on the phenomenon known as *psychological anchoring*,²⁶ it allows for the possibility that people who go to the trouble of rejecting the default will choose a contribution rate close to the default instead of choosing their latent

“individuals whose latent contribution rates are closer to the default are more likely to be affected by the default”

rate. *Anchoring* refers to the tendency for a person asked to choose a numerical value to start with some reference point and then only slightly adjust away from it. In models that focus on opt-out costs, the default does not attract individuals to contribution rates close to but not equal to the default, whereas models that focus on psychological anchoring allow that kind of attraction.

It would be desirable to determine whether the effects of defaults in our empirical setting are driven by opt-out costs or psychological anchoring, but the data do not allow us to distinguish between these two mechanisms. Models that view the default as an anchor allow the default to cause an increase in the fraction of individuals choosing a contribution rate near to but not equal to the default.²⁶ However, these models also feature a countervailing force: Individuals who have that nearby contribution rate as their latent contribution rate are likely to move from their latent contribution rate to the default. On net, the default can lead to either an increase or a decrease in the fraction of individuals choosing that nearby contribution rate. In our real-world data, the default decreases the fraction of individuals choosing nearby contribution rates, but this evidence cannot distinguish between a model of anchoring and a model of opt-out costs because both types of models can predict this empirical pattern.

Policy Implications

Our model applies to many contexts beyond retirement savings. The designer of a smart thermostat can set the default temperature that a home's heating and cooling system targets. The designer of an electronic health record system can set the default number of pills prescribed by a physician for a given patient profile and

medication. The designer of a webpage for charitable contributions can set the default donation amount. The model parameter values that best fit our experimental data are unlikely to be the parameter values that are appropriate when applying the model in other domains. Nonetheless, as we discussed earlier in the article, evidence from a variety of contexts supports the assumptions of the model, suggesting that the structure of the model is indeed applicable in a range of settings.

If a policymaker is trying to increase the mean outcome for a population on some measure, the model provides guidance for selecting a default among ordered options. In general terms, the policymaker should first identify the distribution of latent outcomes. Next, the policymaker should gauge how influential the default is. This information, in turn, should be used to set a default that will pull up the outcomes of many individuals while pulling down the outcomes of few individuals. If the default is weak (that is, if the radius, R , within which the default has an effect, is small), the default that maximizes the mean outcome is likely just above a cluster of popular latent outcomes. If the default is strong (that is, R is large), the default will likely be higher. When F (the value of the adjustment factor for focal, or sticky, options, such as multiples of five) is high, the policymaker should generally avoid placing the default just above latent outcomes that individuals are reluctant to leave (because the default would then pull few individuals up) and should try to place the default just below such latent outcomes (because the default would then pull few individuals down).

For a policymaker to implement this guidance, the ideal approach would be to run an experiment similar to Study 1, featuring a condition with no default (to observe the distribution of latent outcomes) and conditions with defaults (to estimate the strength of the default). If this approach is not feasible, nonexperimental data can be informative. For example, if a company is using a given default at sign-up and finds that few individuals end up with the default option, program managers can infer that the influence of the default is weak and that the distribution of observed outcomes approximates the

distribution of latent outcomes. If managers find that many individuals choose the default option, they can infer that the default's influence is strong. In this latter case, the policymaker would want to push the default to be more extreme so as to shift outcomes in the desired direction.

For additional insight into the likely strength of the default, a policymaker who cannot conduct a study can rely on past research. According to prior work,²² defaults are more effective in domains where individuals are asked to make consumer purchase decisions and less effective in domains where individuals are asked to make pro-environmental decisions, and they are more influential when they communicate the policymaker's recommendation²⁷ or serve as a reference point against which other options are judged²⁸ than when they merely make the default option easy to implement.²³

Our analysis has limitations. The model applies to many settings but not all. For example, in situations where the default influences outcomes primarily because many people are inattentive—that is, they do not notice that a default is being implemented—the model's assumptions regarding the way in which defaults influence outcomes may not be satisfied.²⁹ In these situations, it is less likely to be true that the influence of the default gets weaker as the difference between the default and an individual's latent outcome increases.

This observation highlights a key feature of our experimental setting. The participants made a choice to visit a website for enrolling in a retirement savings plan, so they were paying attention to the decision at hand.²¹ This fact may explain why the default effects we observed are weaker than some other default effects that have been documented previously in studies of retirement savings plans.^{2–6} Perhaps the individuals in our experiment arrived at the website having already thought through the contribution rate they would like to choose and were therefore less susceptible to the default's influence.

In this article, we have not addressed the moral considerations that a policymaker should have in mind when choosing a default. We have

adopted the perspective of a policymaker who is trying to shift outcomes in a particular direction for ethically sound reasons. For example, the policymaker may have strong reasons to believe that psychological biases are causing individuals' choices to deviate systematically from the choices that would maximize their welfare. For another example, the policymaker may wish to shift outcomes because people are making decisions in ways that do not account for the consequences of their choices on others, such as when people consume energy excessively without regard for their contribution to global climate change. Policymakers who are unsure of which outcomes are appropriate should use a different framework for contemplating default selection.²⁶ They should also be careful to avoid subjecting individuals to the risk of significantly negative outcomes.

Our analysis points to some interesting extensions. We considered the choice of a single default for a population of individuals. If those individuals can be divided into easily identifiable subpopulations who have different latent distributions than the full population does and who respond differently to particular defaults (in other words, whose choices are described by different model parameters), it would be possible to tailor a different default for each subpopulation. This line of reasoning can be applied to situations in which the policymaker has a more complex objective than simply shifting mean outcomes of a large population upward or downward. For example, if a policymaker believes that individuals with low incomes have a greater or lesser need for higher retirement plan contribution rates than individuals with high incomes, default policies could be adjusted on the basis of income, with one group's default chosen to increase contribution rates and the other group's default chosen to promote more moderate contribution rates.

As another extension, it would be valuable to consider how a default might change over time. Consider the case of a smart thermostat. To reduce energy consumption at a company, the building managers might initially begin with a default temperature that is only slightly below the temperature that employees would choose

for themselves during winter. After a set time, as the workers habituated to colder temperatures, the managers might lower the default.

Defaults affect the distribution of outcomes in subtle ways. By using our model, policymakers can select defaults for maximal impact.

endnotes

- A. We do not address situations featuring a small number of options (say, five or fewer) in a choice menu. Our model could accommodate such situations, but the structure imposed by our model would be unnecessary. We also do not address situations featuring many unordered options, because our model does not speak to those situations.
- B. The model could also be adjusted for use by policymakers who, for whatever reason, wanted to shift contribution rates downward.
- C. Although the model puts the probability of a default's influence at zero if the latent value is beyond the default's radius of effect (that is, when $|D - L| > R$), in reality, even individuals whose latent contribution rates are very far from the default have some chance of being influenced by the default. We treat the probability as zero for simplicity, on the grounds that it is likely to be much closer to zero than is the case when the latent value is close to the default.
- D. One could make the argument that people whose latent contribution rates are multiples of five might be more likely to be influenced by defaults because they have thought less deeply about their contribution rate choices. However, as we show in the Calibrating the Model section, our calibration exercise indicates that giving F a strictly positive value—0.3—gives the best fit for the data, suggesting that the assumption embedded in our model is the correct one. For additional evidence on the attractiveness of round numbers, see reference 30.
- E. Editors' note to nonscientists: 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.
- F. The model assumes that the effect of the default will be the same for latent values that are an equal distance below or over the default. Additional analyses that test this assumption of a symmetric

effect around the default are described in the Additional Statistical Tests section of the Supplemental Material. The evidence does not contradict the assumption, but the statistical power of the test is low. We view the issue as an interesting one for future research to address.

- G. We do not view these predictions as contradictory to the empirical data because the predicted values fall within the 95% confidence intervals of the corresponding empirical estimates.

author affiliation

Beshears: Harvard Business School and National Bureau of Economic Research. Mason: Voya Financial and Carnegie Mellon University. Benartzi: University of California, Los Angeles. Corresponding author's e-mail: jbeshears@hbs.edu.

author note

We thank Thomas Armstrong and Daniella Listro of Voya Financial for implementing the experiments and preparing the data, and we thank Katherine Milkman for her involvement in the early stages of this research. Andrew

Joung, Seung Hyeong Lee, Predrag Pandiloski, Byron Perpetua, and Kartikeya Vira provided outstanding research assistance. We have benefited from the comments of Thomas Armstrong, Saurabh Bhargava, Lombard Gasbarro, Daniella Listro, Marilyn Morgan, Charles Nelson, Mark Patterson, John Payne, Stephen Shu, Stephen Thomas, and Giovanni Urga. Benartzi is a paid consultant to Voya Financial. Beshears and Benartzi have, at various times, received compensation from and sat on the advisory boards of financial firms. See their websites for complete lists of outside activities. Voya Financial had the opportunity to review the article before public release for the sake of identifying factual inaccuracies, but the authors retained full editorial control of the article contents. The findings and conclusions expressed are solely those of the authors and do not represent the views of Voya Financial, Harvard University, the National Bureau of Economic Research, Carnegie Mellon University, or University of California, Los Angeles.

supplemental material

- <https://behavioralpolicy.org/publications/>
- Method & Analysis

references

- Dubner, S. J. (Host). (2018, December 19). People aren't dumb. The world is hard (No. 340) [Audio podcast episode]. In *Freakonomics radio*. Freakonomics Radio Network. <https://freakonomics.com/podcast/people-arent-dumb-the-world-is-hard-rebroadcast>
- 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>
- Choi, J. J., Laibson, D., Madrian, B. C., & Metrick, A. (2002). Defined contribution pensions: Plan rules, participant decisions, and the path of least resistance. In J. M. Poterba (Ed.), *Tax policy and the economy* (Vol. 16, pp. 67–113). MIT Press.
- Choi, J. J., Laibson, D., & Madrian, B. C. (2004). Plan design and 401(k) savings outcomes. *National Tax Journal*, 57(2, Pt. 1), 275–298. <https://doi.org/10.17310/ntj.2004.2.07>
- Choi, J. J., Laibson, D., Madrian, B. C., & Metrick, A. (2004). For better or for worse: Default effects and 401(k) savings behavior. In D. A. Wise (Ed.), *Perspectives on the economics of aging* (pp. 81–121). University of Chicago Press. <https://doi.org/10.7208/chicago/9780226903286.003.0003>
- Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2008). The importance of default options for retirement saving outcomes: Evidence from the USA. In S. J. Kay & T. Sinha (Eds.), *Lessons from pension reform in the Americas* (pp. 59–87). Oxford University Press.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
- 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>
- Plan Sponsor Council of America. (2018). *60th annual survey of profit sharing and 401(k) plans*.
- Johnson, E. J., & Goldstein, D. (2003, November 21). Do defaults save lives? *Science*, 302(5649), 1338–1339. <https://doi.org/10.1126/science.1091721>
- Abadie, A., & Gay, S. (2006). The impact of presumed consent legislation on cadaveric organ donation: A cross-country study. *Journal of Health Economics*, 25(4), 599–620. <https://doi.org/10.1016/j.jhealeco.2006.01.003>
- Steffel, M., Williams, E. F., & Tannenbaum, D. (2019). Does changing defaults save lives? Effects of presumed consent organ donation policies. *Behavioral Science & Policy*, 5(1), 69–88. <https://doi.org/10.1353/bsp.2019.0005>
- Johnson, E. J., Hershey, J., Meszaros, J., & Kunreuther, H. (1993). Framing, probability distortions, and insurance decisions. *Journal of Risk and Uncertainty*, 7(1), 35–51. <https://doi.org/10.1007/BF01065313>
- Johnson, E. J., Bellman, S., & Lohse, G. L. (2002). Defaults, framing and privacy: Why opting in-opting out. *Marketing Letters*, 13(1), 5–15. <https://doi.org/10.1023/A:1015044207315>
- Levav, J., Heitmann, M., Herrmann, A., & Iyengar, S. (2010). Order in product customization decisions: Evidence from field experiments. *Journal of Political Economy*, 118(2), 274–299. <https://doi.org/10.1086/652463>
- Brown, Z., Johnstone, N., Hascic, I., Vong, L., & Barascud, F. (2013). Testing the effect of defaults on the thermostat settings of OECD employees. *Energy Economics*, 39, 128–134. <https://doi.org/10.1016/j.eneco.2013.04.011>
- Haggag, K., & Paci, G. (2014). Default tips. *American Economic Journal: Applied Economics*, 6(3), 1–19. <https://doi.org/10.1257/app.6.3.1>
- Patel, M. S., Day, S., Small, D. S., Howell, J. T., III, Lautenbach, G. L., Nierman, E. H., & Volpp, K. G. (2014). Using default options within the electronic health record to increase the prescribing of generic-equivalent medications. *Annals of Internal Medicine*, 161(10, Suppl.), S44–S52. <https://doi.org/10.7326/M13-3001>
- Patel, M. S., Day, S. C., Halpern, S. D., Hanson, C. W., Martinez, J. R., Honeywell, S., Jr., & Volpp, K. G. (2016). Generic medication prescription rates after health system-wide redesign of default options within the electronic health record. *JAMA Internal Medicine*, 176(6), 847–848. <https://doi.org/10.1001/jamainternmed.2016.1691>
- Goswami, I., & Urminsky, O. (2016). When should the ask be a nudge? The effect of default amounts on charitable donations. *Journal of Marketing Research*, 53(5), 829–846. <https://doi.org/10.1509/jmr.15.0001>
- Altmann, S., Falk, A., Heidhues, P., Jayaraman, R., & Teirlinck, M. (2019). Defaults and donations: Evidence from a field experiment. *The Review of Economics and Statistics*, 101(5), 808–826. https://doi.org/10.1162/rest_a_00774
- Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, 3(2), 159–186. <https://doi.org/10.1017/bpp.2018.43>
- Carroll, G. D., Choi, J. J., Laibson, D., Madrian, B. C., & Metrick, A. (2009). Optimal defaults and active decisions. *The Quarterly Journal of Economics*, 124(4), 1639–1674. <https://doi.org/10.1162/qjec.2009.124.4.1639>
- 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>
- Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2021). Active choice, implicit defaults, and the incentive to choose. *Organizational Behavior and Human Decision Processes*, 163, 6–16. <https://doi.org/10.1016/j.obhdp.2019.02.001>
- Bernheim, B. D., Fradkin, A., & Popov, I. (2015). The welfare economics of default options in 401(k) plans. *American Economic Review*, 105(9), 2798–2837. <https://doi.org/10.1257/aer.20130907>
- McKenzie, C. R. M., Liersch, M. J., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science*, 17(5), 414–420. <https://doi.org/10.1111/j.1467-9280.2006.01721.x>
- Dinner, I., Johnson, E. J., Goldstein, D. G., & Liu, K. (2011). Partitioning default effects: Why people choose not to choose. *Journal of Experimental Psychology: Applied*, 17(4), 332–341. <https://doi.org/10.1037/a0024354>
- Fowlie, M., Wolfram, C., Spurlock, C. A., Todd-Blick, A., Baylis, P., & Cappers, P. (2020). *Default effects and follow-on behavior: Evidence from an electricity pricing program* (Working Paper 23553). National Bureau of Economic Research. <https://doi.org/10.3386/w23553>
- Pope, D., & Simonsohn, U. (2011). Round numbers as goals: Evidence from baseball, SAT takers, and the lab. *Psychological Science*, 22(1), 71–79. <https://doi.org/10.1177/0956797610391098>



Nudge versus sludge in gambling warning labels: How the effectiveness of a consumer protection measure can be undermined

Philip W. S. Newall, Lukasz Walasek, Elliot A. Ludvig, & Matthew J. Rockloff

abstract

Legal gambling is a large industry in many countries. One way some governments try to protect people from losing more than they can afford is by requiring warning labels on gambling machines and their online equivalents. Prominent labels that make the odds of winning clear serve as *nudges*: They promote a beneficial behavior (such as deciding that the risk of losing money is too high) without interfering with choice (such as by restricting the availability of gambling). However, if gambling operators use labels that are difficult to understand, find, or read, those messages instead hamper decision-making and thus become *sludge*. In this article, we report on new research into whether gambling labels in the world's largest regulated online gambling market (the United Kingdom) are more consistent with nudge or sludge. We found that gambling operators overwhelmingly used sludge strategies when posting required gambling warning labels: For instance, they framed the message using a confusing format, applied a small font size to the text, and placed the warning on obscure help screens. We therefore propose that public policy officials throughout the world establish requirements for the wording and presentation of gambling warning labels to ensure that gamblers are well-informed about the odds they face.

Newall, P. W. S., Walasek, L., Ludvig, E. A., & Rockloff, M. J. (2022). Nudge versus sludge in gambling warning labels: How the effectiveness of a consumer protection measure can be undermined. *Behavioral Science & Policy*, 8(1), 17–23.

“‘sludgy’ labels undercut a player’s ability to choose wisely”

Legal gambling is a large industry in many countries, often in the form of both electronic gambling machines and online gambling.¹ Not surprisingly, many people end up losing more money than they can afford, even when they do not fully meet the criteria for a diagnosis of disordered gambling.^{2,3} In one approach to combating the problem, some governments require gambling operators to post warning labels about the dangers of gambling. In this article, we describe research into the effects of two contrasting approaches to implementing warning labels—which we characterize as delivering a *nudge* or *sludge*—and we report new research into the prevalence of these labeling approaches in the world’s largest regulated online gambling market, the United Kingdom. We also offer recommendations for policymakers who want to reduce gambling’s negative effects on players.

Nudge Versus Sludge

The United Kingdom’s Gambling Commission, which is responsible for regulating the United Kingdom’s online gambling market,⁴ requires that warning information be posted for games such as online roulette and slots.⁵ The *cost of play*, which is the average loss from each gamble given the probability of each outcome and its resulting payoff, is communicated to gamblers via a warning label⁶ that is somewhat akin to the nutrition labels that inform consumers about the contents of their food.⁷ Customized labels go on each game, because the odds differ between games. The probabilities can be calculated precisely for electronic gambling machine games and their online equivalents. The goal of such information is not necessarily to discourage all gambling but to encourage gamblers to understand the risks they are taking when they choose to gamble. Effective warning labels can complement other interventions aimed at reducing the harm caused by gambling.^{8–12}

In behavioral science terms, these warning labels are meant to be *nudges*: interventions that gently prod people to make a decision that is beneficial to them without changing their financial incentives or the freedom to choose.¹³

Policymakers offer consumers information that is relevant to their decision to gamble rather than, say, imposing a rule that restricts how much money they can bet.¹⁴

In reality, though, the labels on online casino games may instead take the form of *sludge*, a term coined by behavioral economist Richard Thaler. Contrary to the intended aims of the government’s warning label mandate, “sludgy” labels undercut a player’s ability to choose wisely. Sludge has received a lot of attention from researchers recently.^{15–18} Most examples come from situations in which consumers are disincentivized from making wise choices by having to fill out unnecessary forms or deal with other sorts of friction.^{15,18} For instance, few people actually obtain manufacturers’ rebates because of the trouble involved in requesting them.¹⁶ As Thaler has pointed out, however, sludge can also have a broader influence, by encouraging “self-defeating behavior such as investing in a deal that is too good to be true.”¹⁶ In this article, we use the definition of sludge in its broad sense.

In the United Kingdom (and elsewhere), two contrasting ways that gambling warning labels can frame cost-of-play figures are *return-to-player* and *house-edge* approaches. The return-to-player format frames the cost of play in terms of the average amount of staked money that is returned as winnings. For example, a U.K. gambler might be told, “This game has an average percentage payout of 90%,” meaning that for every £100 bet, an average of £90 is paid out in prizes and £10 stays with the house. In contrast, the house-edge approach frames the same information in terms of the average amount of money retained by the gambling operator, as in “This game keeps 10% of all

Core Findings

What is the issue?

Although many governments have enacted policies to make the risks of gambling better known to players, these policies are often subverted by operators. This problem is particularly acute in the case of mandatory labeling designed to convey risks and nudge players away from riskier behaviors. Instead of *nudges*, subverted labels can become *sludge*, resulting in impaired decision-making.

How can you act?

Selected recommendations include:

- 1) Ensuring that players are provided with the most useful cost-of-play information possible
- 2) Clearly and unambiguously spelling out the optimal placement of gambling information, as is done for tobacco

Who should take the lead?

Policymakers, legislators, and leaders in entertainment and recreation

money bet on average.” For electronic gambling machine games and their online equivalents, the probabilities can be calculated precisely.

In the next two sections of this article, we cite research by us and others suggesting that the house-edge wording is a nudge whereas the return-to-player label is sludge, and we report on a study we have conducted to examine the prevalence of these two approaches and the visibility of warning labels on U.K. online gambling sites.

Nudges & Sludge in Gambling Warning Labels

Prior research had suggested that framing a gambling warning label in terms of the percentage of money typically kept by the gambling operator (that is, as a house-edge nudge) rather than in terms of the percentage of money returned to players (that is, as a return-to-player sludge message) would result in more gamblers correctly interpreting their odds of winning. In one study, for instance, 25 people who used an electronic gambling machine were presented with a return-to-player message indicating that the average payout was 90% of the money staked. Only six of the 25 selected the correct interpretation in a four-alternative multiple-choice question about the message’s meaning¹⁹—which was that for every £100 bet on the game, about £90 is paid out as prizes. The other participants selected interpretations saying that betting £1 on the game would guarantee a win of 90p, that 90% of people who played would win something, or that the game would give out a prize nine times out of 10.

On the basis of behavioral science research into the importance of framing to the interpretation of messages,^{20–22} we thought that the gambling warning information could be provided in a more understandable way,²³ namely, by speaking in terms of how much money the house keeps—that is, by using house-edge framing. We therefore conducted an experiment comparing the two messaging approaches. In a study with almost 400 participants, 66.5% of gamblers who read a house-edge statement (which explicitly said that the operator keeps an

“regulations do not provide a definition for what ‘easily available’ means”

average of 10% of the money it takes in) selected the correct interpretation of this information in a multiple-choice question, compared with 45.6% of gamblers who were given the equivalent return-to-player statement indicating that an average of 90% of bet money is returned to players.²³ The largest difference was that 32.8% of gamblers given the return-to-player statement incorrectly selected the option of “This game will give out a prize 9 times in 10,” whereas just 10.3% of gamblers given the house-edge statement chose that interpretation. In another study involving 407 gamblers, participants perceived a lower chance of winning when they were provided with house-edge information than when they received equivalent return-to-player information.²³

These divergent interpretations have a significant influence on betting behavior, according to a recent experiment involving more than 2,400 experienced American gamblers.²⁴ Participants were given small amounts of money to keep or gamble with (in the hope of making more); those who chose to gamble played an online slot machine. Those given house-edge warnings were less likely to initiate and continue betting than were participants who read return-to-player labels.

In that study, the warning labels were prominently displayed on the slot machine screens. In the United Kingdom, regulations state that warning information “must be easily available.”¹¹ This phrasing could be interpreted to mean that warning information should be as prominent as a front-of-pack nutrition label.⁷ But because the regulations do not provide a definition for what “easily available” means, gambling operators could conceivably use the ambiguity to avoid displaying the warning so prominently. For this reason, in a new experiment, we examined the visibility of warning labels along with the prevalence of the two cost-of-play formats.

Real-Life Warning Labels

Our study included 363 online roulette games hosted by 26 major gambling operators (see the Supplemental Material for fuller details of our method). The U.K. online gambling sector is vast and growing, so we made an effort to compare one standard product across multiple gambling operators. We chose roulette because it is highly popular in U.K. online gambling²⁵ and because roulette is a standard casino game that is not likely to vary substantively across operators.

A return-to-player statement appeared on 357 of 363 roulette games (98.3%). Not a single house-edge statement appeared. (See note A for information on the labels of the remaining games.) In addition, the labels were difficult to find. To see the return-to-player statements, gamblers had to click an average of 1.28 times ($SD = 0.58$) to navigate away from the main roulette game window. (See note B for more on the statistical terms used in this article.) The screens with the return-to-player statements contained large amounts of text. A later analysis of a random sample of 10 roulette games revealed an average of 2,078 words on the screen (minimum = 875 words, maximum = 3,450 words). Overall, 38.1% of return-to-player statements were found in the first third of the text, 46.8% were found in the middle, and 15.1% were found in the last third of the text. Relative to other text on the screen, 95.5% of return-to-player statements used the smallest font size, and 99.7% (all games but one) used the lowest level of text boldness. Additionally, 16.8% of return-to-player statements used only an acronym to describe the return-to-player information—such as, “RTP is 97.2973%.” Thus, gambling operators overwhelming used sludge rather than a nudge for their required warning label.

These results suggest that the warning labels in current use by gambling operators are less effective than they could be.²⁶ Operators consistently presented low-prominence return-to-player warning labels, even though information that was either more prominent or that used the house-edge framing would likely have been more helpful to consumers. Operators could

have many reasons to choose one format and a specific visibility over another; for instance, perhaps other actors, such as software providers or consultants, made the decision. Regardless of who made the decisions, though, it appears that in the real world of online gambling, what was intended by U.K. regulators to be a nudge has instead turned into sludge.

These findings are relevant to policymakers who set the United Kingdom’s gambling regulations. However, the study has some limitations. Although we surveyed 26 large operators, the roulette games used in the study may have come from the same small pool of industry consultants or software providers. A wider investigation of more games or U.K. operators could be one way to improve on this study, although it could still end up drawing from the same small pool and thus might not meaningfully increase the sample size. Perhaps a better approach would be to investigate a different market. Some research of this sort has been done. For example, a high number of virtual video games across the world now contain gambling-like elements called *loot boxes*,²⁷ which provide randomized rewards, usually for a price. Their use has been linked to problem gambling.²⁸ The Chinese government announced that video game companies would have to publicize loot box odds of winning but gave companies discretion over whether to disclose them in the game or only on their websites.²⁹ A study of Chinese loot box warning labels found that, as with the U.K. online roulette games, most Chinese online video games offering loot boxes provided suboptimal disclosures by displaying this information in hard-to-reach places.³⁰

Policy Recommendations

We know of no previous research examining how gambling operators interpret the regulatory requirement to provide easily available warning labels, and we find it worrisome that so many operators do not make the labels prominent and do not use house-edge framing. We believe that to help reduce the negative consequences of gambling, regulators in the United Kingdom and elsewhere should ensure that players are

sludge

Information designed to undercut a person’s ability to choose wisely



return-to-player label

Framing the cost of play in terms of the average amount of staked money that is returned as winnings



house-edge label

Framing the cost of play in terms of the average amount of money retained by the gambling operator

provided with the most useful cost-of-play information possible so that they can make informed choices when deciding to gamble.

Although prominent house-edge information appears to be more effective than the current return-to-player information given to gamblers, other formats might also turn out to be useful. For instance, a recent study found that a volatility warning, which says that the cost-of-play odds being presented hold true only in the long run, produced beneficial changes in gambling behavior.²⁴ Thus, a label optimized according to the current state of knowledge might read, "This game keeps 10% of all money bet. It takes millions of plays for a gambling game to tend toward its average return. A gambling game will not return a minimum value of prizes in any given period of gambling."

Of course, a label serves no purpose if a player cannot see it. Regulators should clearly and unambiguously spell out the optimal placement of gambling information, as is done for tobacco.³¹ Although cigarette warning labels have changed over the years, regulations in various countries have always insisted on precise and visible positioning of the warning label on the front, sides, or back of the packs. The United Kingdom's loose regulatory guidance on gambling warning label placement has meant that gambling operators often display the required warning labels where gamblers are unlikely to notice or pay attention to them. Lack of clarity on where labels should be positioned is another way that the regulations can encourage self-defeating behavior—that is, they become sludge.

In this article, we have focused on education as a regulatory goal, as is the case with nutrition labeling.⁷ Some regulators, however, may be sufficiently convinced by the evidence on gambling harm^{9,10} to consider strategies beyond straightforward education. With tobacco, behavior change was effected in part by replacing text-based warnings with graphic

"house-edge information appears to be more effective than the current return-to-player information"

warnings that evoke fear.³² Enhancing gambling warning labels with similar fear-based appeals³³ could prove to be even more effective at changing gambling behavior than the warning labels we have discussed, and they are worth considering.³⁴

endnotes

- A. The remaining six games included information that did not deliver return-to-player or house-edge information. Instead, these messages conformed to at least one of the two other allowed categories, which are either "a description of the way the game works and the way in which winners are determined and prizes allocated" or "the probability (likelihood) of winning events occurring."¹¹
- B. Standard deviation is a measure of the amount of variation in a set of values. Approximately two thirds of the observations in a normally distributed data set fall between one standard deviation below the mean and one standard deviation above the mean.

author affiliation

Newall: Central Queensland University and University of Bristol. Walasek and Ludvig: University of Warwick. Rockloff: Central Queensland University. Corresponding author's e-mail: philip.newall@bristol.ac.uk.

supplemental material

- <https://behavioralpolicy.org/publications/>
- Method & Analysis

references

1. The world's biggest gamblers. (2017, February 9). *The Economist*. <https://www.economist.com/blogs/graphicdetail/2017/02/daily-chart-4>
2. Abbott, M. W. (2020). The changing epidemiology of gambling disorder and gambling-related harm: Public health implications. *Public Health*, 184, 41–45. <https://doi.org/10.1016/j.puhe.2020.04.003>
3. Muggleton, N., Parpart, P., Newall, P., Leake, D., Gathergood, J., & Stewart, N. (2021). The association between gambling and financial, social, and health outcomes in big financial data. *Nature Human Behaviour*, 5(3), 319–326. <https://doi.org/10.1038/s41562-020-01045-w>
4. Gambling Commission. (2018). *Review of online gambling*. Internet Archive. <https://web.archive.org/web/20210305232634/http://www.gamblingcommission.gov.uk/PDF/Online-review-March-2018.pdf>
5. Gambling Commission. (2017). *Remote gambling and software technical standards*. Internet Archive. <https://web.archive.org/web/20210111083932/https://www.gamblingcommission.gov.uk/pdf/Remote-gambling-and-software-technical-standards.pdf>
6. Eggert, K. (2004). Truth in gaming: Toward consumer protection in the gambling industry. *Maryland Law Review*, 63(2), 217–286. <https://digitalcommons.law.umaryland.edu/mlr/vol63/iss2/3/>
7. Campos, S., Doxey, J., & Hammond, D. (2011). Nutrition labels on pre-packaged foods: A systematic review. *Public Health Nutrition*, 14(8), 1496–1506. <https://doi.org/10.1017/S1368980010003290>
8. Ettman, C., Abdalla, S. M., & Galea, S. (2018). Applying population health science principles to guide behavioral health policy setting. *Behavioral Science & Policy*, 4(1), 17–23. <https://doi.org/10.1353/bsp.2018.0001>
9. Wardle, H., Reith, G., Langham, E., & Rogers, R. D. (2019). Gambling and public health: We need policy action to prevent harm. *The BMJ*, 365(8198), Article l1807. <https://doi.org/10.1136/bmj.l1807>
10. Browne, M., Langham, E., Rawat, V., Greer, N., Li, E., Rose, J., Rockloff, M., Donaldson, P., Thorne, H., Goodwin, B., Bryden, G., & Best, T. (2016). *Assessing gambling-related harm in Victoria: A public health perspective*. Victorian Responsible Gambling Foundation. <https://responsiblegambling.vic.gov.au/documents/69/Research-report-assessing-gambling-related-harm-in-vic.pdf>
11. Gambling Commission. (2019). *National strategy to reduce gambling harms*. <http://www.reducinggamblingharms.org/asset-library/national-strategy-to-reduce-gambling-harms.pdf>
12. Tucker, J. A. (2018). Fulfilling the promise of choice architecture interventions for addictive behaviors. *Addiction*, 113(7), 1175–1177. <https://doi.org/10.1111/add.14148>
13. Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
14. Noyes, J., & Shepherd, J. (2020). *Gambling review and reform: Towards a new regulatory framework*. Social Media Foundation. <https://www.smf.co.uk/publications/gambling-review-reform/>
15. Sunstein, C. R. (2020). Sludge audits. *Behavioural Public Policy*. Advance online publication. <https://doi.org/10.1017/bpp.2019.32>
16. Thaler, R. H. (2018, August 3). Nudge, not sludge. *Science*, 361(6401), 431. <https://doi.org/10.1126/science.aau9241>
17. Soman, D., Cowen, D., Kannan, N., & Feng, B. (2019). *Seeing sludge: Towards a dashboard to help organizations recognize impedance to end-user decisions and action*. SSRN. <https://ssrn.com/abstract=3460734>
18. Sunstein, C. R. (2018). Sludge and ordeals. *Duke Law Journal*, 68, 1843–1883. <https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=3988&context=dlj>
19. Collins, D., Green, S., d'Ardenne, J., Wardle, H., & Williams, S. (2014). *Understanding of return to player messages: Findings from user testing*. NatCen Social Research.
20. Soll, J. B., Milkman, K. L., & Payne, J. W. (2015). A user's guide to debiasing. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (Vol. II, pp. 924–951). Wiley-Blackwell.
21. Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research*, 15(3), 374–378. <https://doi.org/10.1086/209174>
22. Ungemach, C., Camilleri, A. R., Johnson, E. J., Larrick, R. P., & Weber, E. U. (2018). Translated attributes as choice architecture: Aligning objectives and choices through decision signposts. *Management Science*, 64(5), 2445–2459. <https://doi.org/10.1287/mnsc.2016.2703>
23. Newall, P. W. S., Walasek, L., & Ludvig, E. A. (2020). Equivalent gambling warning labels are perceived differently. *Addiction*, 115(9), 1762–1767. <https://doi.org/10.1111/add.14954>
24. Newall, P. W. S., Byrne, C. A., Russell, A. M. T., & Rockloff, M. J. (2022). House-edge information and a volatility warning lead to reduced gambling expenditure: Potential improvements to return-to-player percentages. *Addictive Behaviors*, 130, Article 107308. <https://doi.org/10.1016/j.addbeh.2022.107308>
25. Gambling Commission. (2019). *Gambling commission industry statistics - May 2019*. Internet Archive. <https://web.archive.org/20201006114707/https://www.gamblingcommission.gov.uk/Docs/Gambling-industry-statistics.xlsx>
26. Page, L. (2021). Disclosure for real humans. *Behavioural Public Policy*, 5(2), 225–237. <https://doi.org/10.1017/bpp.2019.23>
27. Zendle, D., Meyer, R., Cairns, P., Waters, S., & Ballou, N. (2020). The prevalence of loot boxes in mobile and desktop games. *Addiction*, 115(9), 1768–1772. <https://doi.org/10.1111/add.14973>
28. Zendle, D., & Cairns, P. (2018). Video game loot boxes are linked to problem gambling: Results of a large-scale survey. *PLoS ONE*, 13(11), Article e0206767. <https://doi.org/10.1371/journal.pone.0206767>
29. Xiao, L. Y., & Henderson, L. L. (2021). Towards an ethical game design solution to loot boxes: A commentary on King and Delfabbro. *International Journal of Mental Health and Addiction*, 19(1), 177–192. <https://doi.org/10.1007/s11469-019-00164-4>
30. Xiao, L. Y., Henderson, L. L., Yang, Y., & Newall, P. W. S. (2021). Gaming the system: Suboptimal compliance with loot box probability disclosure regulations in China. *Behavioural Public Policy*. Advance online publication. <https://doi.org/10.1017/bpp.2021.23>
31. Hiilamo, H., Crosbie, E., & Glantz, S. A. (2014). The evolution of health warning labels on cigarette packs: The role of precedents, and tobacco industry strategies to block diffusion. *Tobacco Control*, 23(1), Article e2. <https://doi.org/10.1136/tobaccocontrol-2012-050541>

32. Brewer, N. T., Hall, M. G., Noar, S. M., Parada, H., Stein-Seroussi, A., Bach, L. E., Hanley, S., & Ribisl, K. M. (2016). Effect of pictorial cigarette pack warnings on changes in smoking behavior: A randomized clinical trial. *JAMA Internal Medicine*, 176(7), 905–912. <https://doi.org/10.1001/jamainternmed.2016.2621>
33. Witte, K., & Allen, M. (2000). A meta-analysis of fear appeals: Implications for effective public health campaigns. *Health Education & Behavior*, 27(5), 591–615. <https://doi.org/10.1177/109019810002700506>
34. Sillero-Rejon, C., Leonards, U., Munafò, M. R., Hedge, C., Hoek, J., Toll, B., Gove, H., Willis, I., Barry, R., Robinson, A., & Maynard, O. M. (2020). Avoidance of tobacco health warnings? An eye-tracking approach. *Addiction*, 116(1), 126–138. <https://doi.org/10.1111/add.15148>



COVID-19
Vaccine

COVID-19
Vaccine

COVID-19
Vaccine

COVID-19
Vaccine

COVID-19
Vaccine

COVID-19
Vaccine

Encouraging COVID-19 vaccination through behaviorally informed reminders: Results from a national randomized field experiment in Israel

Adi Berliner Senderey, Reut Ohana, Shay Perchik, Ido Erev, & Ran D. Balicer

abstract

Inducing people to get vaccinated is critical for controlling the spread of COVID-19. We explored the effectiveness of two text messaging strategies for encouraging unvaccinated individuals to get their COVID-19 vaccination. One message emphasized social norms to harness people's tendency to act in ways that line up with society's expectations. The other message underscored the personal medical benefits of vaccination. Both messages indicated that the vaccine was reserved for the recipient at a nearby location. Over the course of eight days, the percentage of people who got vaccinated after receiving the medical benefit message was 2.1% higher than the percentage of people who got vaccinated after receiving the social norm message ($p < .001$). Our findings indicate that designing vaccination reminders that highlight the medical benefits of vaccination in addition to the availability of the vaccines can increase vaccination rates.

Berliner Senderey, A., Ohana, R., Perchik, S., Erev, I., & Balicer, R. D. (2022). Encouraging COVID-19 vaccination through behaviorally informed reminders: Results from a national randomized field experiment in Israel. *Behavioral Science & Policy*, 8(1), 25–32.

Finding strategies to boost vaccination rates is essential to controlling the COVID-19 pandemic.^{1–3} In this article, we report on a study we conducted to compare the efficacy of two behavioral science–informed text message–based interventions, which we tested on nearly 800,000 unvaccinated members of Clalit Health Services (CHS), the largest health care provider in Israel. CHS provides primary, specialty, and inpatient care, and its comprehensive data warehouse combines hospital and community medical records.

“emphasizing the social norm could backfire”

messages that underscore the medical benefits of getting vaccinated might be more persuasive.¹³ In our experiment, we set out to test that hypothesis.

Core Findings

What is the issue?

Increasing vaccine uptake against communicable diseases like COVID-19 is critical to managing public health resources and outcomes. To do so, public health authorities have used different messaging strategies. It has been found, however, that messages designed to emphasize the personal medical benefits of vaccination are more effective than those based on social norms and peer pressure.

How can you act?

Selected recommendations include:

- 1) Highlighting the personal medical benefits of getting vaccinated in messaging strategies
- 2) Testing different vaccine messaging strategies' effectiveness to continuously improve outcomes

Who should take the lead?

Leaders and policymakers in public health

Previous studies have demonstrated that carefully worded reminders informing patients that vaccines are reserved specifically for them are effective *nudges*—that is, gentle ways of influencing behavior that neither restrict choice nor significantly change people's economic incentives. In two large multipronged studies, these reminders increased influenza vaccine uptake by an average of 5%.^{4,5} Our experiment builds on this finding, comparing the efficacy of two additional nudges: one that emphasizes social norms and another that focuses on the medical benefits of the vaccine.

Invoking social norms harnesses people's tendencies to act in ways that line up with other people's expectations. This strategy is often effective in encouraging beneficial health-related behaviors.⁶ For example, social norm nudges have outperformed other types of nudges in reducing no-show rates for medical appointments;⁷ they have also increased hand sanitizer use among hospital visitors.⁸ In particular, informing individuals of the most socially acceptable behavior in a given setting or situation has been highly persuasive in motivating them to follow the norm.^{9–12}

In the context of vaccination, however, emphasizing the social norm could backfire. Because vaccines can also protect the unvaccinated through indirect effects, a message that suggests that others have gotten vaccinated can lead to the *free rider* phenomenon, in which people decide to simply rely on the protection others provide rather than to protect themselves.⁶ In situations where the free rider phenomenon might discourage vaccine uptake,

Method

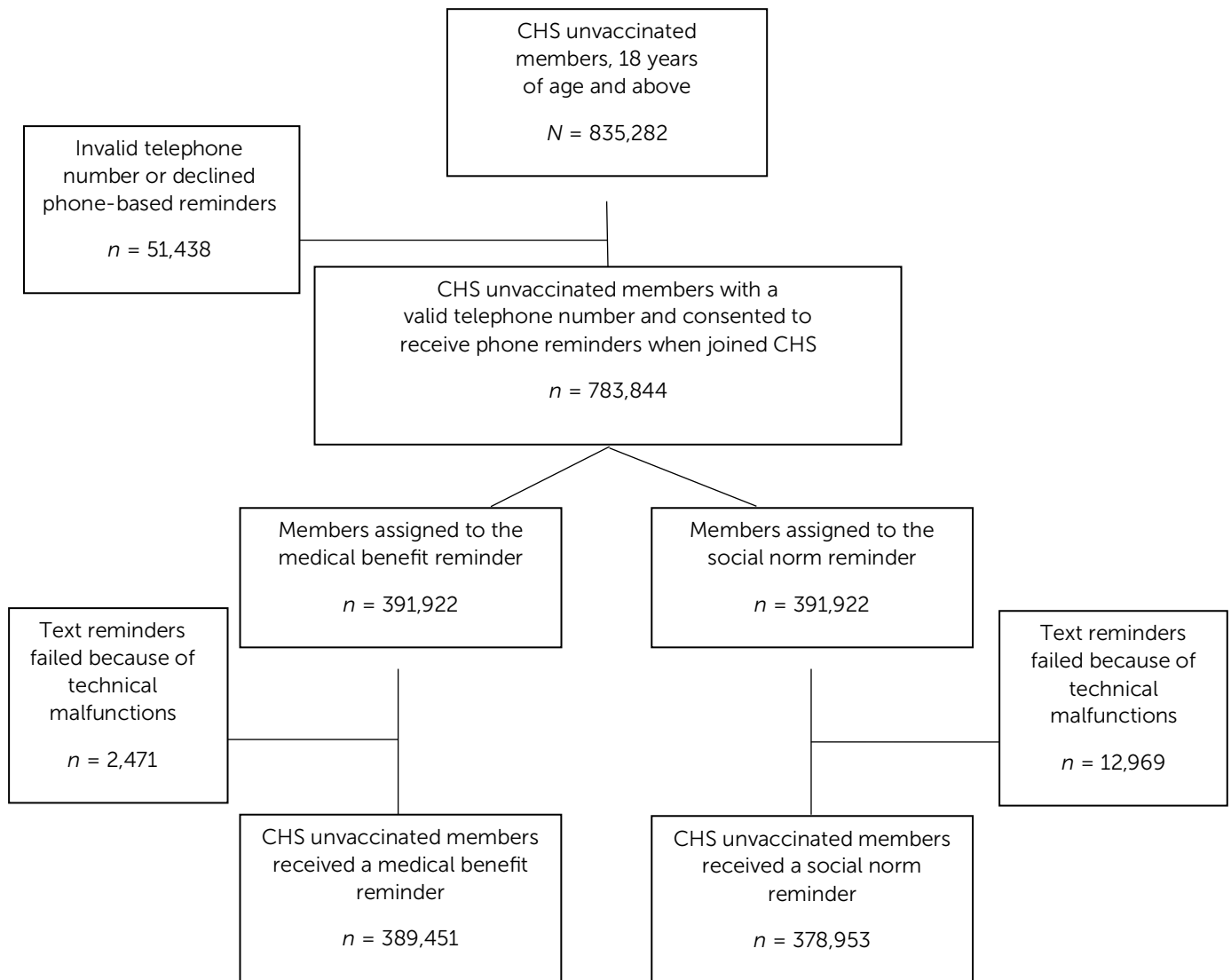
To compare the effects of the social norm and medical benefit nudges we have described, we designed a study in which eligible unvaccinated CHS members were randomly assigned to receive one of two reminders about a week after they received what we called a *baseline* reminder, which merely informed recipients of the vaccine's availability. After another week, we compared the relative effects of the social norm and medical benefit reminders on COVID-19 vaccination rates.

In Israel, COVID-19 vaccination efforts began on December 20, 2020, and by February 4, 2021, vaccines were available to everyone ages 16 years and older. Using CHS's comprehensive health care data warehouse, we identified all unvaccinated members in this age group with a valid cell phone number who, when they joined CHS, had consented to receive text or voice reminders on their cell phones—a total of 783,844 people. See Figure 1 for details of our selection procedure.

On February 8, 2021, all of these individuals were sent the baseline reminder to get vaccinated via the same texting system that CHS uses to send its members reminders of upcoming appointments, notifications of available clinical services, and the like. This message read,

This is a reminder to get vaccinated for COVID-19. This is the quickest way to get back to daily life. The vaccine is available for you at the closest CHS vaccination area [link]. If you'd prefer, you can schedule an appointment [here].

Figure 1. Flowchart showing how Clalit Health Services (CHS) study participants were chosen



Note. If the text reminder failed, members received a voice reminder.

We then randomly assigned each of these individuals to receive one of two additional text message reminders a week later—at 10 a.m. on February 16, 2021—excluding anyone who had scheduled a vaccination appointment by then.

One of these reminders indicated that getting vaccinated was the social norm. It read,

It's time for you to join 3.5 million vaccinated citizens who protected themselves and those they care about. The vaccine is reserved for you today, at the [closest CHS vaccination area] until 21:00. If you'd

prefer, you can schedule an appointment [here].

The other reminder emphasized the medical benefits of the vaccine. It stated,

In a large scale research study conducted by the Clalit Research Institute, the vaccine was found to be effective and reduces 94% of COVID-19-related morbidity! The vaccine is reserved for you today, at the [closest CHS vaccination area] until 21:00. If you'd prefer, you can schedule an appointment [here].

"The medical benefit reminder had an immediate effect"

The number of people who received the social norm and medical benefit reminders by text message was roughly equal—378,953 and 389,451, respectively—but differed slightly because of technical glitches that predominantly affected those assigned to the social norm group. Because of the glitches, an additional 15,440 members who did not receive the text message reminder instead received an identical recorded voice reminder. For our analysis, we focused on the effect of the text message reminders and so excluded those who received the voice reminders from our main analysis, although we did calculate how many of them got vaccinated during the eight-day study period. Including them in the analysis would not have changed the findings appreciably (see Figure S1 in the Supplemental Material for more details).

We made sure that the two groups did not differ clinically, demographically, or in other ways that could confound the results (see Table S1 in the Supplemental Material for details). Approximately half of each group was female, and the average age of participants was 37.5 years for both groups.

We then recorded the percentage of people in each group who received the COVID-19 vaccine each day of the week after the intervention. The data were analyzed using chi-square tests. We considered differences between the groups to be statistically significant when p values from two-sided tests (those allowing for both positive and negative effects of an intervention) were .05 or less. (See note A for more details about the statistics mentioned in this article.)

Results

Figure 2 shows the results. The medical benefit reminder had an immediate effect and was more effective than the social norm reminder. One day after members received our intervention,

the daily vaccination rate was 3.6% among those who received the medical benefit reminder and 3% among those who received the social norm reminder. The advantage of the medical benefit nudge remained clear throughout the study. Over the eight days, the total percentage of people who got vaccinated—the *cumulative vaccination rate*—was 23.8% among members who received the medical benefit reminder and 21.7% for those who received the social norm reminder; that is, the medical benefit reminder increased the vaccination rate 2.1 percentage points more than the social norm reminder did, a statistically significant result ($p < .001$). Both messages improved the vaccination rate over baseline, which was 16.4% just prior the intervention.

We were aware that if individuals in the same household received different intervention messages, the effect of one message could be influenced by the other message, given that these individuals are likely to make vaccination decisions together. We wondered how this dynamic might affect responses to the reminders. We therefore conducted a household-based analysis (see Figure S2 in the Supplemental Material). *Households* were defined as two partners aged 18 years or older living at the same address. (All participants under the age of 18 years received the reminder through their parents.) We included 154,808 households in this analysis, of which 55,329 received heterogeneous reminders (one social norm reminder and one medical benefit reminder). In households in which both members received the medical benefit reminder, the vaccination rate rose to 23.6%, significantly higher than the 21.4% rate in households in which both members received the social norm reminder ($p < .001$), which was in line with our broader findings. In households with mixed interventions, the effects roughly averaged out: 22.7% of members who received the medical benefit reminder got vaccinated and 22.3% of those who received the social norm reminder got vaccinated. The result is consistent with the idea that the medical benefit reminder has a greater effect on vaccination decisions than the social norm reminder does.

3.5m

Vaccinated individuals
in Israel during
February 2021

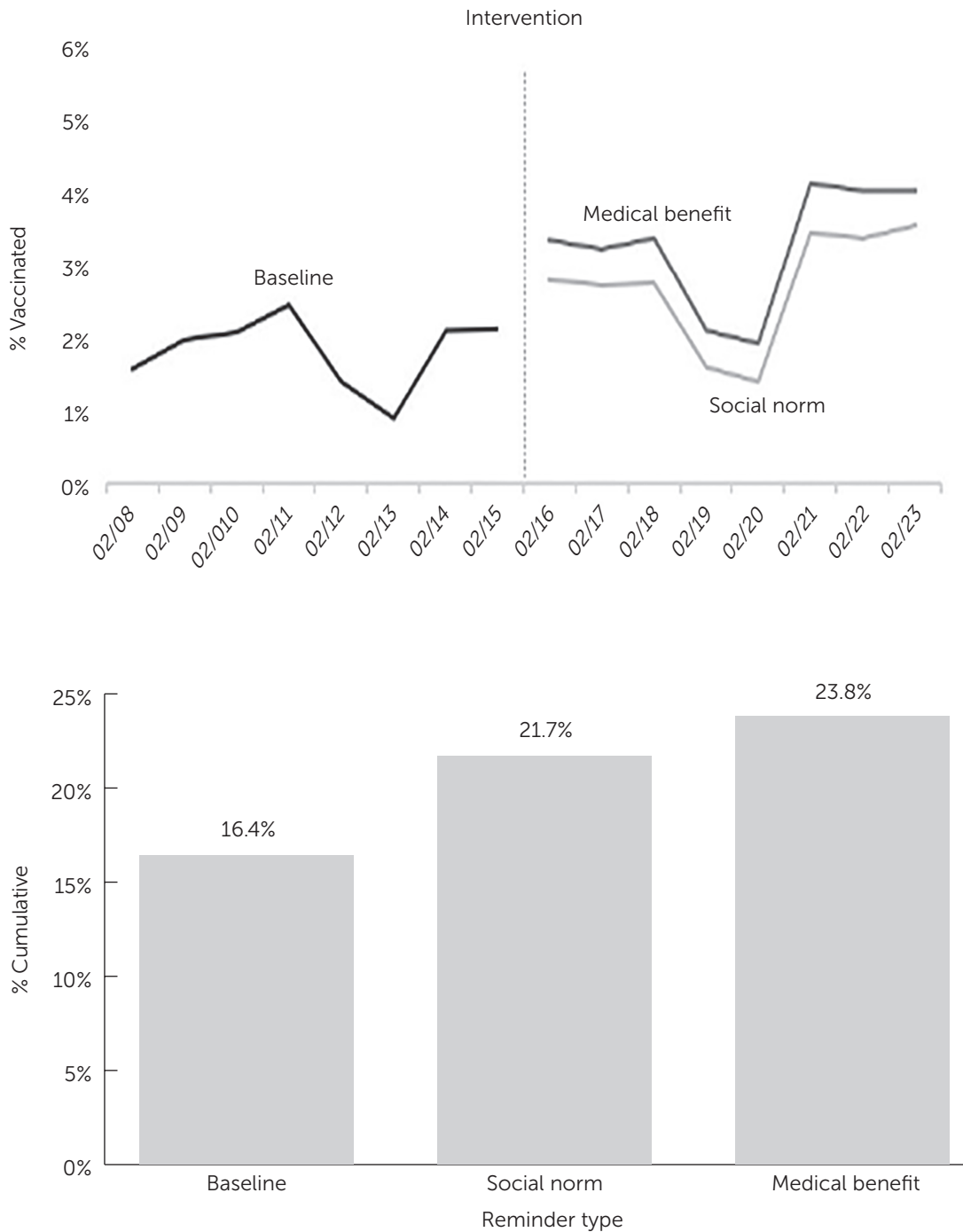
94%

COVID-19 vaccine
efficacy in Israel in
February 2021

2.1%

Increased effectiveness
of medical benefit
messaging over social
norm messaging
for cumulative
vaccine uptake

Figure 2. The percentage of people vaccinated per day before & after the intervention & the cumulative rate



Note. The graphs show that vaccination rates increased more on a daily basis (line graph, top) and cumulatively (bar graph, bottom) after study participants received the medical benefit reminder than after they received the social norm reminder. The dips on February 12–13 and February 19–20 in the line graph reflect lower vaccination rates on the weekends, when some vaccination stations were closed. The increase in the daily vaccination rate after the weekend of February 21—which was similar for both intervention groups—was likely due to the implementation of a national policy that granted vaccinated people access to public facilities.

“text messages that simply inform people that a vaccine is reserved for them can increase vaccination rates”

Conclusion & Policy Implications

Previous research has demonstrated that text messages that simply inform people that a vaccine is reserved for them can increase vaccination rates.^{4,5} In this study, we compared two types of messages that build on the “reserved for you” message and found that an addition emphasizing the medical benefit of a vaccine was more effective than one that focused on the social norm of receiving a vaccine.

Why would the medical benefit message be more effective than the social norm one? One reason may relate to its emphasis on protection. Prior research has shown that reminders stressing personal protection work about as well as those that stress the protection of others in convincing people to get a flu shot.⁵ In our study, the medical benefit reminder underscored both the personal and the societal protection the vaccination provided in its mention of an evidence-based reduction of overall COVID-19-related morbidity (that is, the vaccine reduced the chance of severe disease). By contrast, our social norm reminder did not stress protection so much as apply peer pressure, pointing out that a large number of citizens have already protected themselves and those they care about. As such, our results suggest that when it comes to vaccination, messages that highlight protection or safety, whether individual or societal, may be more convincing than those that rely on people wanting to follow the crowd.

Another reason for the heightened efficacy of the medical benefit reminder could be that it is better at combating procrastination. By the time the intervention took place, COVID-19 vaccines had been available for more than a month.

Some of the individuals who had not yet been vaccinated may have been vaccine skeptics, but at that relatively early stage of the vaccine rollout, many were probably simply procrastinating. The medical benefit reminder may have had an outsized influence on procrastinators because the data supporting a big reduction in morbidity injected some urgency into the vaccination decision.

As Figure 2 shows, a large increase in vaccine uptake occurred about five days after both study interventions, around February 21, 2021. This spike may have stemmed from the national implementation of the Green Pass policy, in which those with vaccine certification (a Green Pass) were granted access to public facilities^{14,15}—a change that would be expected to increase COVID-19 vaccination rates across the nation. The relative advantage of the medical benefit nudge over the social norm nudge did not change, which indicates that this difference is robust.

Overall, our results show that both of the strategies we tested were beneficial but that medical benefit messages like the one we sent would likely be the most effective in future campaigns. Policymakers beyond Israel could use this central finding to potentially improve the effectiveness of their own COVID-19 vaccine promotion programs.

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, with lower values indicating a stronger basis for rejecting the null hypothesis. The analyses in this study were performed with the following software: R version 3.5.3, Python version 3.6, Anaconda version 5.1.0, and tableone version 0.6.6.

author affiliation

Senderey: Clalit Health Services and Technion. Ohana and Perchik: Clalit Health Services. Erev: Technion. Balicer: Clalit Health Services and Ben Gurion University of the Negev. Corresponding author's e-mail: adiber3@gmail.com.

author note

We thank Maya Kotler, Michal Bar Kochva, Yaniv Shimkovich, Ruthi Optovsky, Tami Gorman, Michal Kizelstein, Inbar Zach, Alex Bodic, Adi Hadad, and Anna Krichly for their contributions to the intervention's conceptual framing, its

design, and the implementation of the randomization process. We also thank Sydney Krispin for her editorial support. This study was approved by the Clalit Health Service's institutional review board (0036-21-COM1), which granted a waiver of consent for this research because it was not a clinical trial but rather an organizational initiative to optimize internal policy.

supplemental material

- <https://behavioralpolicy.org/publications/>
- Additional Analysis

references

1. Wood, S., & Schulman, K. (2021, February 18). Beyond politics—Promoting Covid-19 vaccination in the United States. *New England Journal of Medicine*, 384(7), Article e23. <https://doi.org/10.1056/NEJMms2033790>
2. Patel, M. (2021, February 9). Test behavioural nudges to boost COVID immunization. *Nature*, 590(7845), 185. <https://doi.org/10.1038/d41586-021-00329-z>
3. Dai, H., Saccardo, S., Han, M. A., Roh, L., Raja, N., Vangala, S., Modi, H., Pandya, S., Sloyan, M., & Croymans, D. M. (2021, September 16). Behavioural nudges increase COVID-19 vaccinations. *Nature*, 597(7876), 404–409. <https://doi.org/10.1038/s41586-021-03843-2>
4. 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>
5. Milkman, K. L., Gandhi, L., Patel, M. S., Graci, H., Gromet, D. M., Ho, H., Kay, J. S., Lee, T. W., Rothschild, J., Bogard, J. E., Brody, I., Chabris, C. F., Chang, E., Chapman, G. B., Dannals, J. E., Goldstein, N. J., Goren, A., Herschfield, H., Hirsch, A., . . . Duckworth, A. L. (2022). A 680,000-person megastudy of nudges to encourage vaccination in pharmacies. *Proceedings of the National Academy of Sciences, USA*, 119(6), Article e2115126119. <https://doi.org/10.1073/pnas.2115126119>
6. 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>
7. Berliner Senderey, A., Kornitzer, T., Lawrence, G., Zysman, H., Hallak, Y., Ariely, D., & Balicer, R. (2020). It's how you say it: Systematic A/B testing of digital messaging cut hospital no-show rates. *PLoS ONE*, 15, Article e0234817. <https://doi.org/10.1371/journal.pone.0234817>
8. Mobekk, H., & Stokke, L. (2020). Nudges emphasizing social norms increased hospital visitors' hand sanitizer use. *Behavioral Science & Policy*. https://behavioralpolicy.org/journal_issue/covid-19/
9. Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, 35(3), 472–482. <https://doi.org/10.1086/586910>
10. Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2008). Normative social influence is underdetected. *Personality and Social Psychology Bulletin*, 34(7), 913–923. <https://doi.org/10.1177/0146167208316691>
11. Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148, 14–31. <https://doi.org/10.1016/j.jpubeco.2017.02.003>
12. Rogers, T., Goldstein, N. J., & Fox, C. R. (2018). Social mobilization. *Annual Review of Psychology*, 69, 357–381. <https://doi.org/10.1146/annurev-psych-122414-033718>
13. Jung, J. Y., & Mellers, B. A. (2016). American attitudes toward nudges. *Judgment and Decision Making*, 11(1), 62–74.
14. Plonsky, O., Roth, Y., & Erev, I. (2021). Underweighting of rare events in social interactions and its implications to the design of voluntary health applications. *PsyArXiv*. <https://doi.org/10.31234/OSF.IO/9Q3DB>
15. Erev, I., Ingram, P., Raz, O., & Shany, D. (2010). Continuous punishment and the potential of gentle rule enforcement. *Behavioural Processes*, 84(1), 366–371. <https://doi.org/10.1016/j.beproc.2010.01.008>



Ventilator allocations: The effect of mere identifiability

Ilana Ritov & Stephen M. Garcia

abstract

The COVID-19 crisis has raised a dire dilemma among medical professionals. Faced with a shortage of critical equipment and supplies, how do hospital administrators and physicians determine whether to divert resources from one patient to another? Most decision-makers will prioritize saving younger patients over older ones, because older patients generally have a much shorter life expectancy. But emotions, such as those elicited when a patient's name is known and the patient is thereby humanized, can interfere with rational decision-making. At the height of the pandemic, we conducted three studies in which participants were asked to imagine being hospital officials tasked with allocating ventilators under two conditions: when the affected patients were and were not identified by name. Participants were less likely to reassign a ventilator from an older patient to a younger one when the patients had been named than when they had not. These results suggest that decision-makers are more likely to make the efficient choice—the one that should save more years of life—when the individuals affected by the choice remain anonymous. When patients are humanized by being named, less rational and more emotional considerations appear to govern how people choose to distribute lifesaving equipment. Our findings imply that keeping patients anonymous may help facilitate the efficient allocation of scarce medical resources.

Ritov, I., & Garcia, S. M. (2022). Ventilator allocations: The effect of mere identifiability. *Behavioral Science & Policy*, 8(1), 35–44.

Core Findings

What is the issue?

People often do not act optimally when allocating scarce medical resources, especially in times of crisis. When patients' names are known by decision-makers, this can present a barrier to efficiency, potentially costing more lives and/or life years. Humanizing patients, therefore, can sometimes have a collectively detrimental effect on health care outcomes.

How can you act?

Selected recommendations include:

- 1) Establishing priorities for allocating resources
- 2) Requiring uninvolved third parties, not people on the front lines, to allocate scarce resources
- 3) Asking patients directly how they feel about reallocating resources away from them

Who should take the lead?

Health care leaders and policymakers

The COVID-19 pandemic has created a shortage of medical equipment unlike any that the United States has experienced in this century. The limited supply of ventilators and other advanced medical resources creates major roadblocks to providing optimal care for severely ill COVID-19 patients.¹ As a result, professionals in many health care settings have been forced to ration potentially life-saving equipment and have not always done so in a rational manner. In response to this situation, the medical community has sounded the alarm that “hospitals and states urgently need to establish and implement policies that more fairly allocate these scarce resources.”²

But what does “fairness” look like? If, for instance, the supply of ventilators is limited, who is going to receive one? Under a wide range of circumstances, most people will prioritize the health of younger people over older ones.³ This choice, as Geoffrey Goodwin and Justin Landy have put it,

is usually thought to be based on a *years left* argument—all else equal, younger individuals have a greater number of valuable life years ahead and so ought to be prioritized in order to maximize future outcomes . . . and, separately, a *years lived, or fair innings*, argument—younger individuals have not had as much time to live and should be prioritized on the grounds of fairness.³

Yet there are some barriers to maximizing the rational distribution of these resources, or making the “efficient” choice. One of these barriers arises when decision-makers know the names of the patients in question, which tends to humanize them. In this article, we explore how merely identifying a patient who is on a ventilator or who does not have a ventilator but needs one influences the likelihood that a decision-maker will assign the ventilator to the more “deserving” patient. We show that the identifiability of patients weakens the impact of efficiency, or utilitarian, considerations in ventilator assignment and makes allocation decisions more difficult.

Theoretical Background

Identifying individuals makes them more vivid.⁴ Vividness is important because decisions that affect people who seem real are likely to engage people's emotions, whereas decisions about people in the abstract are more likely to be processed in a rational, deliberative mode. The behavioral science literature has shown that merely identifying a target of a decision has the power to change decision-makers' actions in many domains. For example, people are more likely to donate to a cause if one particular beneficiary has been identified than if a group of needy individuals has been presented.^{5–8}

Further, people in general do not like to cause harm, and they particularly do not like to cause harm when those who will be hurt are viewed as individual people rather than as abstract entities. This reluctance to do damage in the face of mere identification has already been demonstrated in other settings. For example, researchers have found that people's support for affirmative action declined more when an individual who would be affected negatively by affirmative action could be pinpointed than when the person was not identified.⁹ Similarly, in the justice system, researchers have determined that people were more lenient in their recommendation for punishment when perpetrators were identified than when they were not.¹⁰

In the case of allocating resources such as ventilators, harm is a real possibility. It is likely that the people making the decision believe (correctly) that the individual removed from the ventilator has a high probability of dying as a result. In this article, we propose that because people are reluctant to harm a specific individual, decision-makers will be more averse to reassigning a ventilator after the patient on the ventilator has been identified by name than when the patient has not been identified.

We further argue that the degree of reluctance to cause harm can differ depending on whether the harm results from action or inaction. Deviating from the current situation—from the status quo—is perceived as an action, whereas accepting the status quo is considered an omission. Our earlier research demonstrates

that people regret a bad outcome of a decision more if the harm was caused by an action than if it resulted from inaction.^{11–13} For hospital workers, moving a ventilator from an older patient to a younger one would be perceived as active harm, because it involves making a change from the existing situation. Not giving the ventilator to the younger candidate, which can be considered the default condition, would be experienced as harm through omission. Thus, when a decision-maker would actively cause harm by moving the ventilator from the older patient to the younger one, knowing the names of the patients would be expected to particularly heighten discomfort over the action and increase the likelihood of deciding that the person already on the ventilator will stay on the ventilator.¹³

Decision-makers may sometimes face a choice between two patients needing a single ventilator when neither person is yet attached to the machine—a situation we describe as having no status quo. In this case, one might expect decision-makers to have no worries about actively causing harm and thus to prefer the efficient option of giving the ventilator to the younger patient. Still, identifying the potential recipients of the ventilator may influence the decision even in this situation. When the patients are identified, emotional reactions—particularly the negative feelings associated with harming by omission the patient who will not receive the ventilator—could pull decision-makers away from efficiency and decrease their preference for the life-maximizing option of giving the ventilator to the younger person.

Method Overview

We tested the effects of mere identification on ventilator allocation by assigning 1,074 participants to three separate studies. The participants were recruited from Amazon Mechanical Turk and paid to participate in this research.

In all three studies, we asked participants to imagine that they were hospital administrators faced with dilemmas that are now occurring around the world because of COVID-19. In Studies 1 and 2, participants were told that

“Identifying individuals makes them more vivid”

their hospital’s intensive care unit (ICU) is completely full, and a new patient has come in with severe respiratory failure. The new patient is a 35-year-old man. He can survive only if he is connected to a ventilator. To provide the ventilator, hospital staffers must disconnect another patient from the potentially lifesaving machine. The other patient, an 83-year-old man, is also experiencing severe respiratory failure. In Studies 1 and 2, the status quo is that the ventilator is currently in use by the older man but could be reassigned to the newly arrived younger man. Study 3 differs from Studies 1 and 2 in that for half of the participants, there is no status quo—the older and younger patients enter the ICU at the same time, and both need a ventilator. Thus, for half of the participants pretending to be hospital administrators in Study 3, making a decision as to which person gets the equipment is not affected by the potential harm of actively taking it away from a gravely ill person.

In all of the studies, we randomly assigned participants to either the *identified* or the *unidentified* condition. In the identified condition, the younger and older patients were named either “Joshua Frey” or “Jack Evers,” with the names counterbalanced across conditions (half of the time the younger patient was Joshua, half of the time he was Jack). In the unidentified condition, no names were mentioned.

Study 1

Method

Study 1 focused on whether identifying patients by name would affect the likelihood that an administrator would make the theoretically efficient decision to reassign a ventilator from an older patient to a younger patient in need of the device. We asked 255 participants (87 women, 168 men) to imagine that they were hospital administrators deciding how to allocate a single ventilator in an ICU. The participants read the following explanatory text:

The ICU in the hospital is completely full, and a new patient comes in with severe respiratory failure. The new patient is 35 years old. The only possibility to treat him is to connect him to a ventilator. To do that you must disconnect from the ventilator another patient, age 83, who also suffers from severe respiratory failure.

Participants were asked to respond yes or no to the question of whether they would disconnect the older patient from the ventilator and give it to the new patient. We predicted that fewer people would reassign the ventilator to the younger patient when the older and younger patients were identified by name than when they were not identified.

Results

Consistent with our hypothesis, the statistical analysis showed that a smaller proportion of the participants said they would disconnect the older patient from the ventilator to help the younger patient when the patients were identified (42.6%) than when they were not identified (55.6%; $\chi^2 = 4.313$, $p = .038$). That is, not identifying the patients led to the more efficient,

or rational, choice. (See Figure 1. Print readers: Color versions of the figures are available online.) For a discussion of the statistical terms used in this article, see note A.

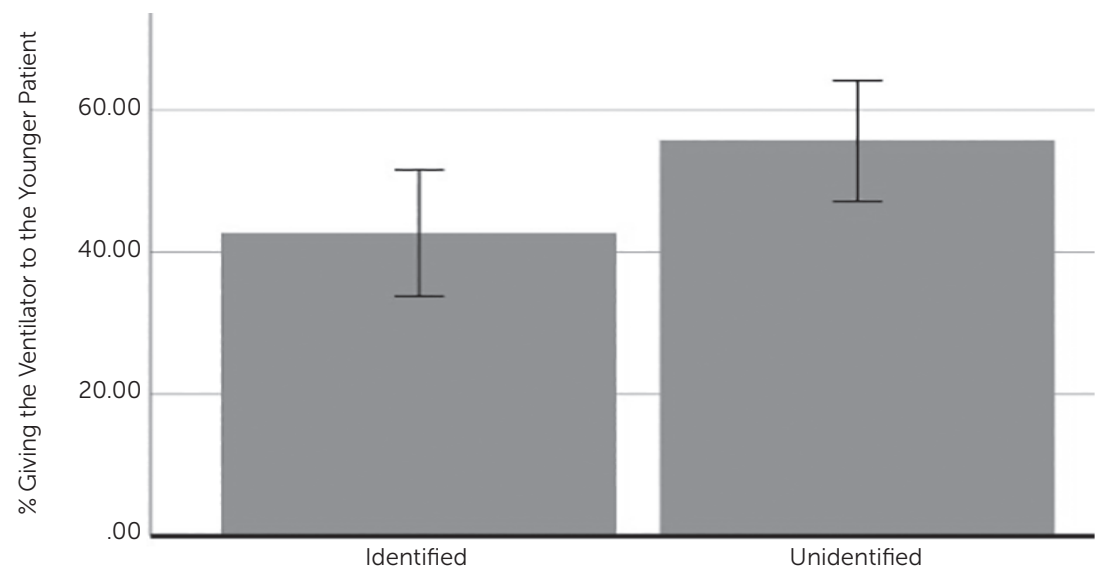
Study 2

Method

In Study 2, we addressed the same question as in Study 1 but also evaluated whether considering hypothetical future patients would produce results that differed from those produced when participants considered unidentified patients in the present. This study involved 350 participants (137 women, 211 men, two unspecified) and three conditions: the two conditions from Study 1 (identified versus unidentified), plus a new one relating to the future.

We randomly assigned participants to the identified, unidentified, and future conditions. In the identified and unidentified conditions, we presented the same scenario we used in Study 1. In the *future* condition, participants read a description of a hypothetical future situation in which an administrator would have to consider reassigning a ventilator from an older patient to

Figure 1. Percentage of participants choosing to disconnect the older patient & give the ventilator to the younger patient (Study 1)



Note. In the identified condition, participants were told the patients' names, whereas in the unidentified condition, patient names were not provided. Error bars reflect 95% confidence intervals.

a younger patient who needed one. This hypothetical situation did not identify the patients by name. Thus, the participants were considering situations relating to unknown people. In that sense, our future condition was similar to the unidentified one.

Participants in all conditions were asked whether they would disconnect the older patient from the ventilator and give it to the new patient. In Study 2, we changed the way we measured the intention to disconnect, now using a seven-point scale ranging from 1 (*definitely no*) to 7 (*definitely yes*). We predicted that because of reluctance to cause harm to named individuals, the participants would be less likely to give the ventilator to the younger patient in the identified condition than in the unidentified and future conditions.

Results

We found a significant difference between the mean ratings of intention to disconnect in the identified and the unidentified conditions ($M = 3.74$ versus $M = 4.28$, $p = .039$), essentially replicating the effect found in Study 1. The higher the means, the more inclined the participants

“participants were considering situations relating to unknown people”

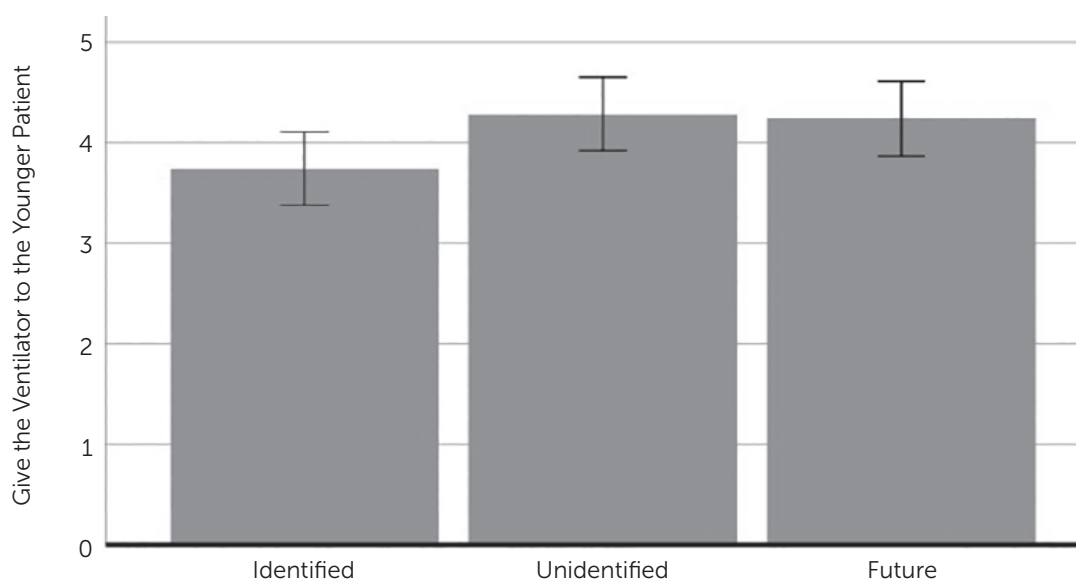
in the role of administrators were to favor the younger life. (See Figure 2.) We also observed a difference between the identified and future conditions that approached significance ($M = 3.74$ versus $M = 4.24$, $p = .060$). As expected, the unidentified and future conditions did not significantly differ ($M = 4.28$ versus $M = 4.24$, $p = .872$). See the Supplemental Material for more details on these analyses.

Study 3

Method

In Study 3, we examined whether decision-makers would respond differently to patient identification in the absence of the influence of an existing status quo—that is, when the older patient was not already on a ventilator. We asked 469 participants (191 women, 277 men, one unspecified) how they would allocate a

Figure 2. Mean ratings of intention to disconnect the older patient & give the ventilator to the younger one (Study 2)



Note. In the identified condition, participants were told the patients' names, whereas in the unidentified condition, patient names were not provided. In the future condition, participants considered the same scenario as a hypothetical future choice between two unnamed patients. Higher numbers indicate a greater tendency to give the ventilator to the younger patient. Error bars reflect 95% confidence intervals.

scarce resource when there was and was not a status quo. Half of the participants received instructions similar to those in Study 1, where the older person had a ventilator to start with. For the other half, the questionnaire posed this dilemma: "You are a hospital administrator when two patients, one young and one elderly, arrive at the ICU at the same time." In this case, then, the older person is not already on the ventilator and no one had to commit the active harm of detaching someone from a ventilator. In other words, this condition negated the predisposition to maintain the status quo.

The *status quo* condition participants were asked whether they would disconnect the older patient from the ventilator and give it to the younger patient. They rated their responses on a 7-point scale ranging from 1 (*definitely no*) to 7 (*definitely yes*). Participants in the *no status quo* condition were asked to identify which patient—the 35-year-old patient or the 83-year-old patient—they would connect to the ventilator. They rated their responses on a 7-point scale ranging from 1 (*definitely the older patient*) to 7 (*definitely the younger patient*). Half of the participants in the status quo and no status quo conditions had the patients identified by name. Study participants thus fell into four groups: identified, no status quo; unidentified, no status quo; identified, status quo; unidentified, status quo.

We hypothesized that when there was no status quo, participants would be more likely to make the rational choice; that is, to give the ventilator to the younger patient. But we also reasoned that identifying the potential recipients by name would heighten the emotional reaction of the decision-makers, especially by eliciting negative feelings about harming the patient who would not receive the ventilator. This reaction would discourage participants from making the life-maximizing choice to give the equipment to the younger patient, who likely had more years to live. Thus, we predicted that identifying the patients by name would reduce life-maximizing efficiencies in both the presence of the status quo, when the decision-maker is deciding whether to reassign the ventilator, and in the

absence of the status quo, when neither participant is on the ventilator.

Results

As in the earlier two studies, our statistical analysis found that participants were less likely to give the ventilator to the younger patient when the two patients were identified by name than when they were unidentified ($M = 4.41$ versus $M = 4.88$). This effect of identifiability on resource allocation was statistically significant, $F(1, 465) = 7.494$, $p = .006$, $\eta_p^2 = .016$.

We also found that whether participants encountered a status quo situation or a no status quo situation had a significant effect on their decisions, $F(1, 465) = 39.603$, $p < .001$, $\eta_p^2 = .078$. As predicted, participants tended less toward giving the ventilator to the younger patient in the status quo situation (in which the older patient was already on the ventilator) than when the two patients showed up at the hospital simultaneously ($M = 4.11$ versus $M = 4.64$).

Additional analyses implied that identifying patients by name influences decision-making in both the presence and the absence of the status quo. (See Figure 3.) Comparing the identified and unidentified means separately for each status quo condition yielded a significant or nearly significant difference in mean intention to give the younger patient the ventilator under each of the two conditions. In the status quo condition, the mean score for the intention to give the ventilator to the younger patient was lower for the identified condition (3.88) than for the unidentified condition (4.35), $t(230) = 1.805$, $p = .072$. Likewise, in the no status quo condition, the mean was 4.94 for the identified condition but 5.39 for the unidentified condition, $t(235) = 2.098$, $p = .037$. See the Supplemental Material for additional statistics.

Aggregate Analysis of the Three Studies

By combining the results of all three studies in an approach known as a *single-paper meta-analysis*,¹⁴ we estimated the impact of patients' identifiability on intentions. Because Study 1 did not use the 7-point scale applied in

Utilitarian decision

Rationalist optimization based on net cost-benefit of assumed outcomes

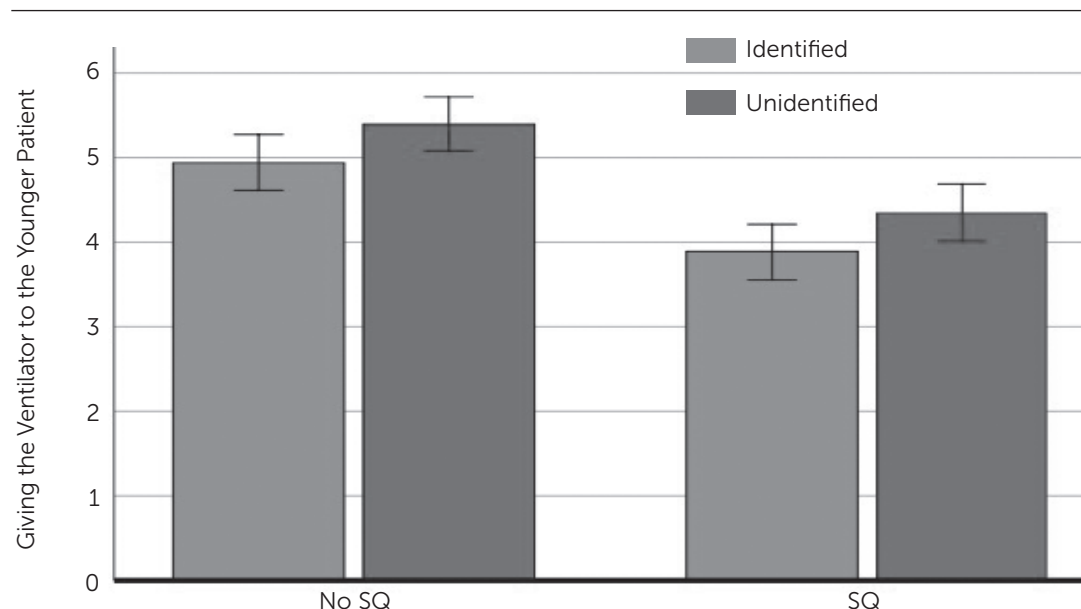
Deontological choice

Following rules that define moral obligations to others

Sacrificial dilemma

A choice between letting several people die and saving those people by sacrificing fewer other people

Figure 3. Mean ratings of intention to give the ventilator to the younger patient (Study 3)



Note: SQ = status quo. In the no status quo condition, neither patient is on the ventilator when the decision regarding who gets the ventilator is made; in the status quo condition, the older patient is already on the ventilator. In the identified condition, participants were told the patients' names, whereas in the unidentified condition, patient names were not provided. Higher numbers indicate greater tendency to give the ventilator to the younger patient. Error bars reflect 95% confidence intervals.

the other two studies, we recoded the data in Studies 2 and 3 so that all three data sets would be comparable. Scores above the midpoint in Studies 2 and 3 were coded as an intention to give the ventilator to the younger patient, and scores below the midpoint were coded as an intention to not give the ventilator to the younger patient. (See the Supplemental Material for more details.) The result was .108, 95% CI [.040, .176]. That is, about 10% more of the decision-makers in this life-or-death situation were inclined to believe that the older patient should have the ventilator instead of the younger one when the patients were identified by name than when they were not. The effect is not large, but it nonetheless seems remarkable, especially when potentially aggregated over hundreds, if not thousands, of lives.

General Discussion

The COVID-19 crisis has revealed the dire dilemma of providing medical services when crucial equipment is in short supply. When resources—such as respiratory ventilators—are scarce, hospital administrators may feel that

the best decisions are those that maximize the common good, such as by saving the most years of patient lives possible. The research presented in this article shows that one factor that may impede people's ability to make rational life-maximizing allocation decisions is the identifiability of the patients. More specifically, we show that people are less likely to reassign a ventilator from an older patient to a younger one when the patients have been identified by name than when they have not (Studies 1 and 2). We also found that identifying patients by name decreases the frequency of choosing the efficient life-maximizing option even when no patient has a ventilator yet—that is, when no status quo has yet been established (Study 3).

Implications for Theory

The dilemma we present in this article shows the tension between two kinds of reasoning that people can engage in simultaneously. On the one hand, decision-makers may feel that the rational, utilitarian decision—what we have called the efficient option—is optimal. Such decisions are based on assumed outcomes, for instance, on the belief that a younger patient on

“identifying by name the persons affected by one’s actions may influence people’s decisions”

a ventilator will live longer than an older patient with the same equipment. On the other hand, people may also feel a need to follow rules, such as the Ten Commandments, that define their moral obligations to others (this is called the *deontological choice*). That is, they may feel it is more ethical to keep alive someone who is already on a ventilator than to remove that ventilator to save someone else. Choosing how to allocate lifesaving ventilators amid a scarcity of resources (when making the efficient decision can feel immoral because it will harm someone) can be viewed as a special case of what has come to be termed a *sacrificial dilemma*. The classic and most studied sacrificial dilemma is the *trolley problem*, in which the decision-maker has to choose whether to sacrifice one person to save several others from a runaway trolley.^{15–20}

Research on the trolley problem suggests that judgments based on emotional reactions to actions rather than to outcomes account for preferring the inefficient option.^{15,16} Although identifiability effects have not been explored in this context, our findings suggest that identifying by name the persons affected by one’s actions may influence people’s decisions. Future research may examine the effect of identifying the victims on the choices people make when faced with trolley-type moral dilemmas.

Implications for Practitioners

On the basis of our research, we offer the following practical advice for hospital officials who play a role in the distribution of scarce medical equipment such as ventilators.

Establish Priorities for Allocating Resources. For hospital employees to make efficient decisions,

some information about the patients in question must be available. For example, age is a legitimate factor to consider: Prioritizing the young over the old is the utilitarian choice. Other factors might include a patient’s preexisting health conditions, whether a patient is a front-line health care worker, and whether a patient is a key decision-maker with responsibility for other people. Establishing and disclosing criteria for allocating resources not only helps avert a dilemma but also adds transparency to the whole allocation process, which will benefit everyone involved: the patients, the medical professionals, the administrators, and the public at large.

Rules are often difficult to craft, yet if criteria are not established, information that may be regarded as irrelevant or even prejudicial—such as one’s social, economic, or national status—may pollute an otherwise fair process. For example, allocation decisions may no longer maximize life if decision-makers are reluctant to reassign a ventilator or any scarce medical equipment from a wealthy old individual to a poor young individual or from a native old individual to a foreign-born young individual. Although age is arguably a relevant factor in such efficient allocations, decision-makers should always beware any irrelevant information that may violate lifesaving goals. That said, we acknowledge that some utilitarian perspectives, however distant from our own, may argue that variables such as wealth are relevant because they signal that someone is more likely to survive and thus deserving of scarce medical equipment. Conversely, some would argue that justice calls for the ventilators to go to historically underserved populations.

Uninvolved Third Parties, Not People on the Front Lines, Should Allocate Scarce Resources.

In cases where protocols have not been established to avert a dilemma, it would be better to have someone other than frontline personnel who are treating the patient make an allocation decision. In this time of COVID-19, medical professionals and staff should learn patients’ names, especially when these patients cannot receive family and friends as visitors. However, if

maximizing life is important, it is more likely that personnel who do not know the patients, such as triage officials, or even impartial computer systems will make the efficient decision. It is important that the frontline worker providing information to the decision-maker not convey any bias, especially with respect to information that might evoke empathy.^{21–24}

Ask the Patient. Finally, a decision-maker can always ask the patient directly about how the patient feels about giving a ventilator to someone else. Indeed, research suggests that people, despite being self-interested, can sometimes make a utilitarian decision even when that decision jeopardizes their own lives.¹⁹ The international press widely reported the story of 72-year-old Italian priest Don Giuseppe Berardelli, who died after choosing to give his ventilator to a younger patient who was in need. Thus, patients themselves may be willing to make a personal sacrifice, even the ultimate one, to maximize someone else's life.

Even though the supply of ventilators has increased since the start of the pandemic, similar dilemmas around allocating life-saving resources to patients may well transpire again, perhaps during a future pandemic. We hope that our work leads to the creation of protocols that will ease the decision-making process in these difficult situations. In addition, we believe that beyond demonstrating the effect of identifiability on life-and-death decisions in the era of COVID-19, our research contributes more generally to the understanding of identifiability as a barrier to efficient choice.

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. F tests and t tests are *parametric*: They make some assumptions about the characteristics of a population, such as that the compared groups have an equal variance on a compared factor. In cases where these assumptions are violated, researchers make some adjustments in their calculations to take into account dissimilar variances across groups. 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 the chance question, researchers consider how much effect a variable has on the statistical results, using measures such as η_p^2 (partial eta squared); η_p^2 values of .01, .06, and .14 typically indicate small, medium, and large effect sizes, respectively. 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. 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.

author affiliation

Ritov: Hebrew University. Garcia: University of California, Davis. Corresponding author's e-mail: ilana.ritov@huji.ac.il.

supplemental material

- <https://behavioralpolicy.org/publications/>
- Method & Analysis

references

1. Boodman, E. (2020, April 14). Who gets the last ventilator? Facing the coronavirus, a hospital ponders the unthinkable. *STAT News*. <https://www.statnews.com/2020/04/14/coronavirus-who-gets-last-ventilator-hospital-ponders-unthinkable/>
2. White, D. B., & Lo, B. (2020). A framework for rationing ventilators and critical care beds during the COVID-19 pandemic. *JAMA*, 323(18), 1773–1774. <https://doi.org/10.1001/jama.2020.5046>
3. Goodwin, G. P., & Landy, J. F. (2014). Valuing different human lives. *Journal of Experimental Psychology: General*, 143(2), 778–803. <https://doi.org/10.1037/a0032796>
4. Small, D. A., Loewenstein, G., & Slovic, P. (2007). Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes*, 102(2), 143–153. <https://doi.org/10.1016/j.obhdp.2006.01.005>
5. 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>
6. Kogut, T., & Ritov, I. (2005). The “identified victim” effect: An identified group, or just a single individual? *Journal of Behavioral Decision Making*, 18(3), 157–167. <https://doi.org/10.1002/bdm.492>
7. Kogut, T., & Ritov, I. (2005). The singularity effect of identified victims in separate and joint evaluations. *Organizational Behavior and Human Decision Processes*, 97(2), 106–116. <https://doi.org/10.1016/j.obhdp.2005.02.003>
8. Slovic, P. (2010). If I look at the mass I will never act: Psychic numbing and genocide. In S. Roeser (Ed.), *Emotions and risky technologies* (pp. 37–59). Springer. https://doi.org/10.1007/978-90-481-8647-1_3
9. Ritov, I., & Zamir, E. (2014). Affirmative action and other group tradeoff policies: Identifiability of those adversely affected. *Organizational Behavior and Human Decision Processes*, 125(1), 50–60. <https://doi.org/10.1016/j.obhdp.2014.04.002>
10. Lewinsohn-Zamir, D., Ritov, I., & Kogut, T. (2016). Law and identifiability. *Indiana Law Journal*, 92(2), Article 3. <https://www.repository.law.indiana.edu/ilj/vol92/iss2/3>
11. Ritov, I., & Baron, J. (1995). Outcome knowledge, regret, and omission bias. *Organizational Behavior and Human Decision Processes*, 64(2), 119–127. <https://doi.org/10.1006/obhd.1995.1094>
12. Baron, J., & Ritov, I. (2009). Protected values and omission bias as deontological judgments. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 50, pp. 133–167). Elsevier. [https://doi.org/10.1016/S0079-7421\(08\)00404-0](https://doi.org/10.1016/S0079-7421(08)00404-0)
13. Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: Omission bias and ambiguity. *Journal of Behavioral Decision Making*, 3(4), 263–277. <https://doi.org/10.1002/bdm.3960030404>
14. McShane, B. B., & Böckenholt, U. (2017). Single-paper meta-analysis: Benefits for study summary, theory testing, and replicability. *Journal of Consumer Research*, 43(6), 1048–1063. <https://doi.org/10.1093/jcr/ucw085>
15. Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and Social Psychology Review*, 17(3), 273–292. <https://doi.org/10.1177/1088868313495594>
16. Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001, September 14). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105–2108. <https://doi.org/10.1126/science.1062872>
17. Aguilar, P., Brüssino, S., & Fernández-Dols, J. M. (2013). Psychological distance increases uncompromising consequentialism. *Journal of Experimental Social Psychology*, 49(3), 449–452. <https://doi.org/10.1016/j.jesp.2013.01.002>
18. Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology*, 104(2), 216–235. <https://doi.org/10.1037/a0031021>
19. Huang, K., Greene, J. D., & Bazerman, M. (2019). Veil-of-ignorance reasoning favors the greater good. *Proceedings of the National Academy of Sciences, USA*, 116(48), 23989–23995. <https://doi.org/10.1073/pnas.1910125116>
20. Barak-Corren, N., Tsay, C.-J., Cushman, F., & Bazerman, M. H. (2018). If you’re going to do wrong, at least do it right: Considering two moral dilemmas at the same time promotes moral consistency. *Management Science*, 64(4), 1528–1540. <https://doi.org/10.1287/mnsc.2016.2659>
21. Bloom, P. (2017). *Against empathy: The case for rational compassion*. Random House.
22. Bloom, P. (2017). Empathy and its discontents. *Trends in Cognitive Sciences*, 21(1), 24–31. <https://doi.org/10.1016/j.tics.2016.11.004>
23. Decety, J., & Cowell, J. M. (2014). The complex relation between morality and empathy. *Trends in Cognitive Sciences*, 18(7), 337–339. <https://doi.org/10.1016/j.tics.2014.04.008>
24. Bazerman, M., Bernhard, R., Greene, D., Huang, K., & Barak-Corren, N. (2020, April 29). How should we allocate scarce medical resources? *Harvard Business Review*. <https://hbr.org/2020/04/how-should-we-allocate-scarce-medical-resources>



Encouraging employees' active feedback & participation when rolling out major changes

Elizabeth A. Hood & Jean M. Bartunek

abstract

When managers and other leaders of organizational change (*change agents*) introduce and implement major changes, responses from other members of the organization can fall along both passive–active and positive–negative dimensions. Change agents usually treat positive (*approving*) responses as good and negative (*disapproving*) responses as bad. They often ignore the passive–active dimension—that is, the degree of energy with which organizational members affected by change (*change recipients*) respond to the initiative. We suggest that change agents instead focus more on this passive–active dimension and work to elicit active responses to change even when these responses are negative, because active responses can lead to valuable improvements in the initiative. We provide three recommendations for assisting organizational leaders in encouraging and learning from the active feedback of change recipients.

Hood, E. A., & Bartunek, J. M. (2022). Encouraging employees' active feedback & participation when rolling out major changes. *Behavioral Science & Policy*, 8(1), 47–58.

Organizational change is inevitable: To improve a business's functioning, people at all levels must adapt to altered environments, embrace novel tools and techniques, and experiment with new processes. Such transitions and transformations are considered successful when they are implemented well (such that they are understood by organization members and endure after their initial rollout) and achieve their stated objectives. Today, the pace of organizational change is accelerating around the world for reasons as far ranging as technological and scientific developments, globalization, climate change, and the ongoing coronavirus pandemic.^{1,2} Yet research suggests that 50% to 80% of change initiatives fail.^{3–5}

One of the most important factors influencing the success or failure of these initiatives is how the people we call *change recipients* (typically the employees whose work processes are altered) feel about and act on proposed and actual changes. Also important is how constructively the *change agents* (often the managers who direct these initiatives) address these responses. An illustration comes from Avinor, a private aviation company administered by the Norwegian government.⁶ In 2003, the company announced an urgent need to cut costs. Initially, employees accepted this news, both because civil aviation around the world was in poor economic straits and because management created processes to engage employees across levels in developing a plan to reduce expenses. But when the final plans—which included extensive layoffs and closing facilities—became public, support for the effort collapsed, because employees realized that the change initiative did not incorporate their feedback. Unions, which had initially supported Avinor's changes, pushed back, challenging the plans openly and requesting an investigation of leadership's numbers. By 2006, although management had implemented some 80% of the company's proposed changes, the remaining 20% (which were meant to produce the most significant cost savings) had not been accomplished.

Managers and other change agents often blame failed initiatives on change recipients'

resistance.⁷ It is true that employees who are upset about a new initiative sometimes try to subvert it.⁸ It turns out, however, that even negative responses (as long as they are in good faith) can help a change succeed if the change agents make use of the respondents' concerns to improve the proposal.⁹ In fact, Avinor's experience features a common detail in stories of unsuccessful organizational change: When management fails to truly engage with the feedback and perspectives of all members of an organization, the lack of engagement can doom even plans with core goals that have strong support across a company.

Unfortunately, inattention to recipients' responses when major changes are initially proposed or rolled out can not only undermine the effectiveness of the initiatives but also, at times, put the organization or the people it serves at risk. Employees, after all, often know the details of how changes can affect work processes and can identify ways that the initiative can be improved as well as aspects that might seriously damage the organization. Elizabeth A. Hood, the first author of this article, observed this dynamic firsthand when she worked for a company in which a manager altered processes so that the company could claim more work was completed for a large client than was actually the case. This unethical behavior was initially hidden from Hood. When she recognized it, she felt that, given the company's culture, she could not say anything to her manager or others about ethical issues. She chose to ignore the directive and proceed in an ethical manner. Her response eventually led to her removal from that particular project and strengthened her desire to leave the company. In this case, management's ignorance of Hood's reaction ultimately harmed the company.

Another example comes from the recent Boeing 737 MAX crisis, which had devastating consequences. Boeing management had changed plans and decided to speed up the deployment of the 737 MAX series airplanes; later, when two of these planes crashed because of a system failure, investigations revealed that employees had already identified the problem but their perspectives had not been taken into account.¹⁰

Core Findings

What is the issue?

Driving successful organizational changes is never easy. *Change agents*—leaders and managers—usually focus on the *valence* of organizational members' responses to change, that is, whether they are negative or positive. However, for changes to be successful, change agents need to consider *activation*: how active or passive these change recipients' responses are as well.

How can you act?

Selected recommendations include:

- 1) Seeking to understand not only whether employees approve or disapprove of a change but also what drives their response
- 2) Focusing primarily on encouraging active responses—whether positive or negative—to change

Who should take the lead?

Organizational leaders and managers

Lives were lost because the company failed to give sufficient weight to employees' knowledge and experience.

Next, we explore some factors that influence how change recipients make sense of initiatives—that is, how they assess the value and consequences of the initiatives—and how they feel and act on the basis of their understanding. Then we outline how scholars categorize employee reactions in terms of both how activated they are (how much energy employees display) and how positive or negative those reactions are. We argue that highly activated responses are at least as valuable to the organization and sometimes even more important than highly positive responses. Finally, we provide recommendations to help managers foster active responses to change initiatives.

How Employees Make Sense of Change

The recipients of organizational change evaluate new initiatives by how the changes may affect them personally and the organization broadly.^{9,11} They often identify the possible gains and losses from the initiative to determine how they want to respond.^{9,11} Individuals also evaluate a new proposal on the basis of what they know about the people implementing this change. For instance, they consider the managers' apparent attitudes toward them as employees.¹² When a manager and an employee have a good relationship, the employee tends to be more supportive of a new initiative. Individuals also think about change initiatives in light of how well aligned managers' stated changes are with what actually occurs when management introduces new processes.¹³

In addition, research shows that people who help shape, carry out, and provide insights into a change initiative typically form more positive perceptions of that change.¹⁴ Thus, managers' involving recipients in the change process can greatly influence the success of the change initiative.

Further, studies suggest that as a change progresses, individuals' responses typically

“as a change progresses, individuals' responses typically shift”

shift.¹⁵ Some individuals shift from negative to positive evaluations, while others shift from positive to negative.¹⁵

Types of Recipient Responses to Change

Organizational researcher Shaul Oreg and his colleagues⁹ have categorized change recipient responses on the basis of *activation* and *valence*. That is, individuals can respond to a change with passive or active behaviors and with positive or negative feelings. (See Table 1.)

Low-activation or *passive* responses refer to reactions that involve relatively little energy, regardless of whether the person's feelings are positive or negative. For instance, individuals may simply not express their ideas about a given change.⁹ Absenteeism and low engagement are also examples of passive responses.^{9,16,17} By contrast, *active* responses are those that can be described as comparatively high energy, regardless of their valence. These responses, whether in words or in deeds, provide considerable feedback to managers and other change agents. This can lead to long-term positive consequences, even when the feedback causes some initial delays stemming from the change initiative's being adapted or altered.⁹ With respect to valence, meanwhile, change recipients might embrace (feel positively toward) or reject (feel negatively about) a new initiative for any number of reasons.

Next, we outline what combinations of passive–active and positive–negative responses to change look like.⁹ To help translate past research so that managers and others can put its conclusions into practice, we describe the categories of change responses and how these responses apply in specific instances of change. We also describe some field studies that suggest contexts in which the types of responses might

occur. These illustrations are important because change recipients' responses do not occur in a vacuum, and case studies help illuminate the factors that can contribute to them. Finally, we include advice for ways that managers can respond constructively to the varied responses an initiative may receive.

Of course, organizational leaders do not always have time to consider employee responses before making changes—as when the COVID-19 pandemic suddenly forced employers to have their staff work from home. In this article, we focus on circumstances in which change recipients' responses may have an important impact on the change.

Change Acceptance

In some cases, change recipients may approve of a proposed new initiative and do little to interrogate or engage with the change agents putting forward the proposal. In these cases, the change recipient may feel calm, relaxed, and contented.⁹ For example, when a manager recommends a new approach that seems to be a good idea to all involved, the employees may say they support it without raising any questions.

This scenario might seem, on the surface, to be ideal. Indeed, when the change is simple and straightforward—not requiring initiative on the part of the change recipients—this turn of events creates no problems. However, when

organizational members passively accept a change, they often do not take steps to explore what the initiative will mean in practice. As a result, they do not raise questions, give feedback, or engage in troubleshooting that would help the new process be implemented effectively.^{9,18} In these cases, the initial rollout may appear to go smoothly and quickly, but unexpected problems may arise during the change's full implementation.^{9,19,20}

Illustrative Field Study. In 2007, researchers John McAvoy and Tom Butler described an attempt at organizational change within a team of programmers.²¹ The change was intended as a new approach to software development, in which team members would collect stories of software user experiences as a way of receiving feedback. Team members were initially quite enthusiastic about this idea.

However, the collection and use of user stories was more complicated than it first appeared. After reviewing some of the collected user stories, team members realized that sharing the views and opinions found in the user stories might create conflict within their team, forcing the group to discuss topics about which team members had significantly differing opinions.²¹ The team members became reluctant to discuss user stories. Some individuals even placed obstacles in the way of applying insights from user stories—for instance, by arguing that the

Table 1. Matrix of responses to change initiatives

	Passive behavior	Active behavior
Positive feelings	Recipients agree with the change without interrogating or discussing it; they feel calm, relaxed, and content. Although seemingly beneficial, this response can signal that change recipients have not had an opportunity to think through the practical consequences (and potential pitfalls) of a change.	Recipients proactively engage with the change, offering feedback and asking questions; they feel excited, elated, and enthusiastic. Feedback in these cases is often extremely valuable, as the recipients of the change are eager to see it succeed and want to actively facilitate the process.
Negative feelings	Recipients disengage from the change process; they feel despair, sadness, and helplessness. This response may signal employees' skepticism that managers will seriously attend to or engage with their perspectives and concerns. This reaction can predict problems not only for a change initiative but also within the organizational culture or for a given employee's relationship with the organization as a whole.	Recipients resist the change; they feel stressed, angry, and upset. Provided the recipients of change have the best interests of the organization at heart, this response—although confrontational—signals an opportunity for management to learn more about how to modify approaches to the change in ways that can significantly benefit the organization.

Note. Based on Figure 2 in "An Affect-Based Model of Recipients' Responses to Organizational Change Events," by S. Oreg, J. M. Bartunek, G. Lee, and B. Do, 2018, *Academy of Management Review*, 43(1), p. 70 (<https://doi.org/10.5465/amr.2014.0335>). Copyright 2018 by Academy of Management.

stories needed more documentation than the project manager had initially thought necessary. Once team members started to encounter and acknowledge such difficulties, their commitment to the approach “effectively disappeared,” McAvoy and Butler noted.²¹ The team stated that this process “was not worth the effort,” even though, in theory, they continued to support the collection of user stories.²¹ Team members eventually reverted to their prior approach to software development, and the change initiative failed.

Change Proactivity

A positive, energetic approach to change is also possible.⁹ In these cases, employees and other change recipients support the change and foster its implementation. These individuals are generally eager to be engaged and provide feedback, which may lead to *breakthrough ideas*, or insights that emerge from asking important questions and applying the answers to carry out the change initiative.^{9,22} Such proactive responses promote successful change within the organization while also fostering interaction and collaboration between the people rolling out the change and those receiving it. People who react proactively in this manner are likely to express support for organizational change, defend it against attacks, and help develop it. Emotions related to change proactivity include excitement, elation, and enthusiasm.⁹

Illustrative Field Study. Two organizational researchers, Lotte Lüscher and Marianne Lewis, conducted a field study of a change initiative in the Lego Group.²³ The company was undergoing a comprehensive restructuring that would create self-managing teams at every level; this change would essentially eliminate the distinction between lower- and middle-level managers. The middle managers involved would need to find a way to understand and accept major changes in their managerial roles for the change to meet its objectives and thereby enable the new management structure to effectively replace the older one.

The Lego Company engaged Lüscher as a consultant. She conducted “sparring” sessions with the middle managers to help them work

“Disengaged employees may feel despair, sadness, and helplessness”

through and troubleshoot scenarios with this new management structure both before and after its introduction.²³ In these sessions, the managers identified challenges they experienced and examined dilemmas they could not easily solve. For instance, managers were concerned about being responsible for the results of teams that they needed to allow to function more independently than they had in the past.²³ Lüscher also helped them learn to allow leaders of the self-managing teams to work more autonomously while still requiring those teams to provide timely reports that enabled planning.²³ These exercises allowed the managers to develop strategies to address problems inherent in the organizational change rather than being paralyzed by these challenges.²³ Lüscher was helpful (and probably crucial) in encouraging the managers to be actively engaged in the process. As a result, the organization implemented the initiative smoothly.

Change Disengagement

In contrast to a passive, positive responder, who simply goes along with a change, a passive, negative responder quietly disengages from the change process. Disengaged individuals do not exhibit strong negative attitudes and may even appear to accept a change without actually doing so. Rather than complain or push back against the new proposal, they may take such actions as absenting themselves from discussions about a change, making errors, and dragging their feet at the implementation stage.^{9,24,25} Disengaged employees may feel despair, sadness, and helplessness.⁹ Compared with other employees, they tend to be more cynical²⁶ and feel more negatively about their job.^{26,27} Hood’s example of her own experience represents a type of change disengagement. Hood did not challenge the manager: Given her prior experiences with the company, she believed that an active response would not

succeed. She simply did not go along with the unethical change.

As is true with people who accept change passively, the limited feedback provided by people who are disengaged may lead managers to mistakenly assume that implementation of a change will go well. However, implementation may be difficult if it depends on the participation of these disengaged individuals.⁹ Change agents are often unable to determine the reasons for disengagement and so lose opportunities to improve the change they are trying to lead.

Illustrative Field Study. Timo Vuori and Quy Huy documented Nokia's attempts several years ago to develop touch screen phones to compete with Apple's iPhone.²⁸ Their study illustrates how managerial actions might lead to employee disengagement.

Their report revealed that prior to Apple's introduction of the iPhone, Nokia's top managers had information that the new product would feature a touch screen, which Nokia phones did not have.²⁸ Upper management told middle managers that touch screens needed to be included in Nokia phones. However, upper management did not share the strategic rationale for doing so, nor did they adequately explain to middle managers the importance of touch screens to the future success of Nokia phones. The top managers did not have strong technical competence themselves and were more concerned about external competitors and shareholders than they were about their middle managers or their subordinates, whom they assumed could be pressured into accomplishing what was wanted.

Thus, middle managers were focused on accomplishing the technical tasks without understanding their overall purpose.²⁸ These middle managers were afraid of upper management, because top managers at the company had a history of putting pressure on and responding very aggressively toward the middle managers. Vuori and Huy described the atmosphere as being one of fear.²⁸ One middle manager, they noted, "was typically open and clear, but in certain meetings [with higher-level

leaders], he became very quiet and when he spoke his voice was shaking."²⁸ Because of the culture of fear, the middle managers also were afraid of their colleagues in other units, so they never criticized others' ideas, wary of what would happen to them in response. Even when Nokia's middle managers knew that incorporating touch screens was proving difficult, they did not share this negative information with their top managers. Rather, they made over-optimistic promises that were ultimately unrealistic.²⁸ "Fearing top managers' immediate negative reactions, they remained silent or filtered information," Vuori and Hoy noted.²⁸ These responses gave top managers a distorted view of how Nokia was doing in response to the iPhone. The result of this dysfunction was that the company did not take the necessary steps to successfully deploy a touch screen phone.²⁸ Nokia's phones declined in quality and usability, product introductions ran late, the CEO was eventually dismissed, and these failings ultimately contributed to the decline and downfall of the company.

Change Resistance

Change recipients who have an active, negative response purposely set in motion forces that interfere with the successful implementation of change.^{9,29} The emotions associated with change resistance include stress, anger, and upset.^{9,30,31}

Change resistance can initially seem undesirable, and in the short term, it often is. Active resistance is likely to slow down the change process and perhaps disrupt it entirely.^{9,32} It can also lead to reduced commitment to the organization and to lowered perceptions of organizational effectiveness.³³ However, when resistance is not actively destructive, it can result in employees providing helpful feedback about the reasons a change may be problematic for its recipients. Some members resist organizational change because they desire to see the organization succeed and have legitimate concerns about a change. Therefore, in the long term, resistance may—if recognized and addressed—be helpful to the initiative.⁹ In this way, some forms of resistance can potentially result in proposals that foster the success of the

organizational change process.⁹ When change agents fail to recognize that change resistance is sometimes helpful, they squander an opportunity to use feedback from employees to adapt change processes in a constructive way.³⁴

Resistance to organizational change benefits an initiative most during the planning stage,⁹ allowing change agents to learn about potential problems and modify their approach to avoid them. Change resistance is much more problematic once implementation begins. Thus, finding ways to obtain feedback, including negative feedback, from change recipients early on is a powerful policy.

Illustrative Example and Field Study. Many timely examples of change resistance have come about because of policies linked to the COVID-19 pandemic. As the highly contagious novel coronavirus spread in early 2020, companies had to adapt to the pandemic in part by switching to primarily remote work arrangements. As vaccines became available and virus cases declined, leadership at some organizations began to announce transitions back to prepandemic modes of working, for example, by requiring employees to come back into the office five days a week. However, in certain cases, employees have pushed back against such initiatives, citing concerns about the continued spread of the virus or the benefits of more flexible work arrangements.³⁵ The responses of these individuals often illustrate thoughtful resistance to change initiatives.

An interesting field study of active resistance focused on members of an organization that is not a workplace but rather a religious organization, the Catholic Church. Betzaluz Gutierrez and her colleagues studied a case of resistance that was a response to a lack of organizational change that church members believed was crucial.³⁶ In 2002, the *Boston Globe* began to publish revelations that some Catholic priests had abused children and that some bishops had subsequently covered up these incidents. Many Catholics felt pain, anger, betrayal, and shock at this news, and church attendance dropped precipitously. Some people who considered themselves “concerned Catholics” formed a

group called Voice of the Faithful (VOTF). This group became a vehicle that gave lay Catholics a means of responding to the scandal, so as to take part in the church’s governance and alter “the passivity (among lay Catholics) that members came to see as a root cause of the abuse and cover-up.”³⁶ This group grew to about 20,000 members within a few months. Partly in response to VOTF’s prodding, Catholic bishops in the United States did eventually establish a charter for the protection of children. The members of VOTF felt that taking action “enabled them to become part of the solution in an embattled institution,” Gutierrez and her colleagues noted.³⁶

Recommendations for Managers

On the basis of our earlier discussion, we suggest three key ways that change agents, building on their understanding of responses to change, can best ensure a change initiative’s success:

Recommendation 1: Seek to Understand Not Only Whether Employees Approve or Disapprove of a Change but Also What Drives Their Response

The examples and case studies together reveal that for managers to fully benefit from responses to a change initiative, they need to understand how employees feel about the change, why employees feel that way, and what kind of response these feelings evoke. As our examples have indicated, employee feelings drive each of the different types of recipient responses to change.

Determining these drivers of responses to change may require thinking about an individual’s past behavior and context. Perhaps a quirk of personality is the reason behind an individual’s tendency to behave passively in many situations. Or an individual may have personal reasons for being withdrawn at work that have nothing to do with the organization or the change the organization is trying to make. Speaking directly with change recipients is essential if change agents are to get a better sense of where the recipients’ feelings are coming from; it is especially important when recipients’ responses are

change agents

Leaders and directors of organizational change

change recipients

Employees whose work processes are altered

50% to 80%

Change initiatives that ultimately fail

In Brief: How Managers Can Foster the Success of Change Initiatives

Recommendation 1: *Seek to understand not only whether employees approve or disapprove of a change but also what drives their response.*

- Determine the reasons for employees' responses to change so you may respond appropriately. Ask employees what they think and why.
- Think about each employee's response in light of larger circumstances—for instance, is this response typical of that employee's behavior, or is it symptomatic of larger cultural norms at the organization that need to be addressed?
- To practice perspective taking, consider yourself a recipient of a recipient's responses to change and explore the reasons for your own response. Think about what you might do differently on the basis of this experience to elicit active constructive feedback from employees.

Recommendation 2: *Strive to incorporate recipient feedback rather than spending energy trying to convert recipients' evaluations from negative to positive.*

- When seeking out the opinions of employees about a change, do not automatically reject or push against negative feedback.
- Seriously consider the merits of the employees' feedback (for example, with recommendation 1 in mind, think about whether the feedback comes from a desire to support the organization and its goals overall).
- Use employee feedback to make appropriate judgments about the change: Are there things that can and should be altered in the new initiative? Has the initiative's purpose been clearly communicated? How can the organization prepare for some of the problems that employees foresee?

Recommendation 3: *Focus on encouraging active responses—whether positive or negative—to change.*

- Make it clear in communication that you are seeking out and value diverse opinions on a new initiative.
- Develop structures and processes for employees to provide feedback about change. This step is particularly valuable after announcing a change and before it has begun in earnest, as employee responses can have a greater impact at this stage.
- Engage employees in implementing the change: People who feel they are shaping an initiative will be more likely to support its execution.

negative. When employees respond negatively to a change initiative because of their legitimate concerns about its viability and perhaps its ethics, managers need to weigh and act on their feedback.

Managers can use the feedback they receive as an opportunity to practice perspective taking. When they receive feedback, they essentially become recipients of a proposal for change and can reflect on this experience further. They can pay attention to how it feels to be on the receiving end of new ideas for how to do things. For example, does an employee's critical feedback make them feel upset? That response may provoke their own change resistance. What would it take to direct the energy of their response productively and engage with the employee's ideas? By asking such questions, managers can practice identifying their own emotional responses and consider how they might communicate current plans and future proposals to engage others in a productive, active way.

Recommendation 2: Strive to Incorporate Recipient Feedback Rather Than Spending Energy Trying to Convert Recipients' Evaluations From Negative to Positive

Several case studies we outlined have demonstrated the damaging consequences of failing to encourage and take into account feedback from change recipients. In these cases, leadership would have been wise to amplify and attend to negative voices rather than dismissing or disagreeing with them. This is because negative feedback can provide leaders with information and direction to improve a change initiative, and this information may be coming from people in the organization better positioned than the leaders to know about the problems a change initiative might be heading for. When managers seek opinions on a change, they gain the opportunity to identify and address valid concerns about an initiative and avoid potential pitfalls. Thus, rather than attempting to convert individuals to a positive viewpoint, managers should seek ways to gather a variety of different opinions and incorporate all types of recipient feedback into the change initiative.

In addition, change agents who attend to such feedback may discover that a negative response reflects a failure in their own communication. For instance, the change recipients may not understand an initiative's purpose or importance to the organization. To counter disengagement, then, managers in these cases could share their reasons for a given change. Although efforts such as increasing communication and modifying plans on the basis of employee feedback often take time, they can, when done well, lead to greater employee commitment to implementing the change and, ultimately, to greater success in reaching the goals of the. Eventually, a change initiative is likely to need buy-in from the people who will actually implement the changes. This end will be most easily accomplished when managers acknowledge their employees' initial concerns and encourage them to participate in shaping how an initiative is carried out.

Recommendation 3: Focus Primarily on Encouraging Active Responses—Whether Positive or Negative—to Change

Given that active responses to change, whether positive or negative, provide more feedback than passive ones, change agents should consider approaches that assist in shifting employee responses from passivity toward activity. For example, managers can hold meetings in which both those who evaluate a change positively and those who view it negatively are invited to air their perspectives—a move that signals interest in the expression of multiple viewpoints rather than a preference for positive feedback. This kind of managerial response would likely have fostered more successful change at Nokia.

Managers can sometimes inadvertently contribute to passivity. For instance, they may convey that they want employees to simply accept a proposal without taking time to scrutinize it. In some cases, managers may have a larger hurdle to overcome, as when employees are afraid to give voice to their concerns about a change initiative.²⁸ Managers should consider ways to reduce this fear. A key factor is allowing individuals to share their honest evaluations without fear of management retaliation.^{10,37} Managers could address this issue

“Managers can sometimes inadvertently contribute to passivity”

by communicating about the change in ways that give employees permission to disagree, providing opportunities to ask questions, creating a forum for offering feedback, or all of the above. Managers should also evaluate whether they can do more to express interest in employee concerns and viewpoints. When managers both provide vehicles for response and demonstrate that they take this feedback seriously, employees feel more motivated to share their ideas and feelings.

Managers can also tailor strategies for eliciting active engagement based on the feelings they perceive as driving responses, as suggested in recommendation 1. Further, when individuals need additional guidance to more fully understand a change, managers might engage in strategies similar to those described in the Lego case study.²³ A consultant who truly understands organizational processes can be particularly helpful in these cases. Giving employees a chance to talk through, spar with, and test-drive scenarios related to new processes and workflows will provide them with the opportunity to offer feedback and ask questions in a safe space.

As general guidance, we recommend that managers develop a structure and process for employees to provide feedback. For instance, regular meetings in which managers seek honest feedback from recipients could serve as a forum in which active responses to change are encouraged and shared. Once these structures and processes are in place, employees will be more likely to become actively engaged in the change initiative. Over time, as employees begin to feel that they have an influence on the change initiative, buy-in should increase and promote the successful implementation of the change. Therefore, managers should seek greater activation and engagement of recipients during a change initiative.

Summary & Conclusion

We have argued that the passive–active dimension of responses to change is likely more important than the positive–negative dimension. When change recipients respond with energy, it is easier for change agents to get feedback that supports both the likely success of the change at hand and the health of the organization as a whole. The field studies we have presented have shown that each type of employee response to change comes within a context—such as a culture of fear or an atmosphere in which only positive responses are

encouraged—that managers help to create. If managers are to foster change effectively, they need to understand how change recipients are responding to the managers’ actions and then react constructively to those responses. We hope that the examples and suggestions we have provided in this article will help steer the course of change in many organizations.

author affiliation

Hood and Bartunek: Boston College. Corresponding author’s email: hoodeb@bc.edu.

references

1. Baker McKenzie. (2017). *Global Employer Forum 2017: FutureWorks*. https://www.bakermckenzie.com/-/media/files/expertise/employment/the-global-employer-forum_2017.pdf?la=en
2. Innovation Vista Editorial Staff. (2020, November 15). Change is faster than ever, but will never again be this slow. *Innovation Vista*. <https://innovationvista.com/digital/change-fast-never-this-slow/>
3. Burke, W. W. (2011). A perspective on the field of organization development and change: The Zeigarnik effect. *The Journal of Applied Behavioral Science*, 47(2), 143–167. <https://doi.org/10.1177/0021886310388161>
4. Jorgensen, H.-H., Bruehl, O., & Franke, N. (2014). *Making change work... while the work keeps changing: How change architects lead and manage organizational change*. IBM Corporation. <https://www-935.ibm.com/services/us/gbs/bus/html/gbs-making-change-work.html>
5. Strebler, P. (1996). Why do employees resist change? *Harvard Business Review*, 74(3), 86–92. <https://hbr.org/1996/05/why-do-employees-resist-change>
6. Loftquist, E. A. (2011). Doomed to fail: A case study of change implementation collapse in the Norwegian civil aviation industry. *Journal of Change Management*, 11(2), 223–243. <https://doi.org/10.1080/14697017.2010.527853>
7. Battilana, J., & Casciaro, T. (2013). Overcoming resistance to organizational change: Strong ties and affective cooptation. *Management Science*, 59(4), 819–836. <https://doi.org/10.1287/mnsc.1120.1583>
8. Ramirez, I. L., & Bartunek, J. (1989). The multiple realities and experiences of internal organisation development consultation in health care. *Journal of Organizational Change Management*, 2(1), 40–57. <https://doi.org/10.1108/09534818910134086>
9. Oreg, S., Bartunek, J. M., Lee, G., & Do, B. (2018). An affect-based model of recipients' responses to organizational change events. *Academy of Management Review*, 43(1), 65–86. <https://doi.org/10.5465/amr.2014.0335>
10. Beer, M. (2021, January 27). It's time employees had a voice. *Mike Beer*. <https://www.beermichael.com/blog-1/its-time-employees-had-a-voice>
11. Bartunek, J. M., Rousseau, D. M., Rudolph, J. W., & DePalma, J. A. (2006). On the receiving end: Sensemaking, emotion, and assessments of an organizational change initiated by others. *The Journal of Applied Behavioral Science*, 42(2), 182–206. <https://doi.org/10.1177/0021886305285455>
12. Balogun, J., Bartunek, J. M., & Do, B. (2015). Senior managers' sensemaking and responses to strategic change. *Organization Science*, 26(4), 960–979. <https://doi.org/10.1287/orsc.2015.0985>
13. Bartunek, J. M., Lacey, C. A., & Wood, D. R. (1992). Social cognition in organizational change: An insider-outsider approach. *The Journal of Applied Behavioral Science*, 28(2), 204–223. <https://doi.org/10.1177/0021886392282004>
14. Bartunek, J. M., Greenberg, D. N., & Davidson, B. (1999). Consistent and inconsistent impacts of a teacher-led empowerment initiative in a federation of schools. *The Journal of Applied Behavioral Science*, 35(4), 457–478. <https://doi.org/10.1177/0021886399354007>
15. Jansen, K. J., Shipp, A. J., & Michael, J. H. (2016). Champions, converts, doubters, and defectors: The impact of shifting perceptions on momentum for change. *Personnel Psychology*, 69(3), 673–707. <https://doi.org/10.1111/peps.12120>
16. Farrell, D. (2017). Exit, voice, loyalty, and neglect as responses to job dissatisfaction: A multidimensional scaling study. *Academy of Management Journal*, 26(4), 596–607. <https://doi.org/10.5465/255909>
17. Reichers, A. E., Wanous, J. P., & Austin, J. T. (1997). Understanding and managing cynicism about organizational change. *Academy of Management Perspectives*, 11(1), 48–59. <https://doi.org/10.5465/AME.1997.9707100659>
18. March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2(1), 71–87. <https://doi.org/10.1287/orsc.2.1.71>
19. Amiot, C. E., Terry, D. J., Jimmieson, N. L., & Callan, V. J. (2006). A longitudinal investigation of coping processes during a merger: Implications for job satisfaction and organizational identification. *Journal of Management*, 32(4), 552–574. <https://doi.org/10.1177/0149206306287542>
20. Holt, D. T., Armenakis, A. A., Feild, H. S., & Harris, S. G. (2007). Readiness for organizational change: The systematic development of a scale. *The Journal of Applied Behavioral Science*, 43(2), 232–255. <https://doi.org/10.1177/0021886306295295>
21. McAvoy, J., & Butler, T. (2007). The impact of the Abilene paradox on double-loop learning in an agile team. *Information and Software Technology*, 49(6), 552–563. <https://doi.org/10.1016/j.infsof.2007.02.012>
22. Berger, W. (2014). *A more beautiful question: The power of inquiry to spark breakthrough ideas*. Bloomsbury.
23. Lüscher, L. S., & Lewis, M. W. (2008). Organizational change and managerial sensemaking: Working through paradox. *Academy of Management Journal*, 51(2), 221–240. <https://doi.org/10.5465/amj.2008.31767217>
24. Oreg, S. (2006). Personality, context, and resistance to organizational change. *European Journal of Work and Organizational Psychology*, 15(1), 73–101. <https://doi.org/10.1080/13594320500451247>
25. Pinder, C. C., & Harlos, K. P. (2001). Employee silence: Quiescence and acquiescence as responses to perceived injustice. *Research in Personnel and Human Resources Management*, 20, 331–370. [https://doi.org/10.1016/S0742-7301\(01\)20007-3](https://doi.org/10.1016/S0742-7301(01)20007-3)
26. Reichers, A. E., Wanous, J. P., & Austin, J. T. (1997). Understanding and managing cynicism about organizational change. *Academy of Management Perspectives*, 11(1), 48–59. <https://doi.org/10.5465/ame.1997.9707100659>
27. Stanley, D. J., Meyer, J. P., & Topolnitsky, L. (2005). Employee cynicism and resistance to organizational change. *Journal of Business and Psychology*, 19(4), 429–459. <https://doi.org/10.1007/s10869-005-4518-2>
28. Vuori, T. O., & Huy, Q. N. (2016). Distributed attention and shared emotions in the innovation process: How Nokia lost the smartphone battle. *Administrative Science Quarterly*, 61(1), 9–51. <https://doi.org/10.1177/0001839215606951>
29. Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, 74(4), 967–984. <https://doi.org/10.1037/0022-3514.74.4.967>
30. Fugate, M., Kinicki, A. J., & Scheck, C. L. (2002). Coping with an organizational merger over four stages. *Personnel Psychology*, 55(4), 905–928. <https://doi.org/10.1111/j.1744-6570.2002.tb00134.x>

31. Bordia, P., Jones, E., Gallois, C., Callan, V. J., & DiFonzo, N. (2006). Management are aliens! Rumors and stress during organizational change. *Group & Organization Management*, 31(5), 601–621. <https://doi.org/10.1177/1059601106286880>
32. Schweiger, S., Stouten, H., & Bleijenbergh, I. L. (2018). A system dynamics model of resistance to organizational change: The role of participatory strategies. *Systems Research and Behavioral Science*, 35(6), 658–674. <https://doi.org/10.1002/sres.2509>
33. Jones, S. L., & Van de Ven, A. H. (2016). The changing nature of change resistance: An examination of the moderating impact of time. *The Journal of Applied Behavioral Science*, 52(4), 482–506. <https://doi.org/10.1177/0021886316671409>
34. Ford, J. D., Ford, L. W., & D'Amelio, A. (2008). Resistance to change: The rest of the story. *Academy of Management Review*, 33(2), 362–377. <https://doi.org/10.5465/amr.2008.31193235>
35. Allen, J., & O'Keeffe, D. (2021, July 6). Your employees stepped up in a crisis. What happens when it's over? *Harvard Business Review*. <https://hbr.org/2021/07/your-employees-stepped-up-in-a-crisis-what-happens-when-its-over>
36. Gutierrez, B., Howard-Grenville, J., & Scully, M. A. (2017). The faithful rise up: Split identification and an unlikely change effort. *Academy of Management Journal*, 53(4), 673–699. <https://doi.org/10.5465/AMJ.2010.52814362>
37. Burris, E. R., & Sohn, W. (2021). Creating a culture of voice. *Behavioral Science & Policy*, 7(1), 57–68.



“Braking” bad: How managers can respond to employee misbehavior

Erin L. Frey, Gabrielle S. Adams, Evan A. Bruno, & James R. Detert

abstract

Employee misbehavior can be defined as transgressions that go beyond unintentional mistakes but do not rise to the level of criminal offenses. Managers are often given substantial discretion over how to handle such behavior, but they may be unsure about what their response should be or unaware of the extent to which others will care about their response. We offer a framework to help managers respond to misbehavior, particularly when firing the offender is not an option. We identify types of formal and informal responses that not only deter future offenses but also help to restore perceptions of justice within the organization. We also provide guidance on how managers should select and communicate these responses to other employees. Finally, we highlight two supplementary actions that managers can consider to restore perceptions of justice: *victim restitution* (that is, providing compensation to or otherwise helping to assuage the distress of the wronged party) and *offender reintegration* (that is, helping the wrongdoer get back to work within the organization).

Frey, E. L., Adams, G. S., Bruno, E. A., & Detert, J. R. (2022). “Braking” bad: How managers can respond to employee misbehavior. *Behavioral Science & Policy*, 8(1), 61–77.

Core Findings

What is the issue?

Employee misbehavior frequently lies somewhere between unintentional mistakes and criminal misconduct. Managers therefore face confusion about the appropriate course of action when employees intentionally and seriously misbehave, especially when this conduct does not rise to the level where employment can or should be terminated. Using appropriate informal and formal sanctions, which support both victim restitution and offender reintegration, can restore a sense of organizational justice.

How can you act?

Selected recommendations include:

- 1) Considering actions that support victims, such as compensatory justice and apologies
- 2) Supplementing punitive responses with developmental responses when misbehavior could be partly unintentional

Who should take the lead?

Managers and organizational leaders

In 2010, an employee of the Australian bank Macquarie was taking part in a live TV broadcast about interest rates. The interview went viral when viewers noticed that another Macquarie banker was clearly browsing pornography in the background.¹ Employee wrongdoing—whether petty theft, insubordination, or something as disturbing as the employee browsing pornography while at work—is unfortunately common. When such behavior occurs, managers—that is, people with authority and supervisory duties in an organization—are frequently called on to respond. Like all people, managers generally want justice to be served but also want to be seen as fair and compassionate. Further, managers are highly motivated to deter future offenses, because employee transgressions can be costly: US companies lose \$50 billion annually from employee theft alone.² Even so, managers may be unsure of how best to respond to wrongdoing or may not even realize that people might want them to respond.

Wrongdoing can vary in intentionality, severity, and legality. In Table 1, we sort various employee transgressions into three categories: unintentional mistakes, serious misbehavior, and criminal misconduct. What constitutes an appropriate managerial response largely depends on the category into which an employee's behavior falls. For example, if a bank teller embezzles money, this action violates the law. A manager's appropriate response would be to fire the employee and report the embezzlement to a law enforcement agency, such as the FBI in the United States or the Serious Fraud Office in the United Kingdom, for further action. In contrast, if a teller miscounts the cash in a drawer, then a manager's appropriate response to this unintentional mistake would be to provide an intervention intended to instruct the offender so the mistake would not be made again.³

Between unintentional mistakes and criminal misconduct lies *serious misbehavior*, which, as Yoav Vardi and Yoash Wiener have put it, includes "any intentional action . . . that violates core organizational or societal norms."⁴ Although serious misbehavior may not be easily defined, scholars generally distinguish between three types: *interpersonal misbehavior* (such

as harassment, bullying, or incivility), which harms individuals in an organization;⁵ *group misbehavior* (such as discrimination and social undermining), which harms individuals and impairs the organization or a subset of people within the organization;^{6,7} and *organizational misbehavior* (such as theft, vandalism, absenteeism, and substance use), which harms the organization as a whole rather than specific individuals.

Researchers have studied many types of serious misbehavior, including workplace deviance,⁶ wrongdoing and misconduct,⁴ employee mistreatment,⁸ counterproductive work behavior,⁹ transgressions,¹⁰ violations,¹¹ and offenses.¹² The common thread in all of these concepts is the understanding that serious misbehavior breaks social or cultural norms and harms an organization or its members.¹³

Responding to these cases can be challenging. Because serious misbehavior falls below the threshold of criminal misconduct, responses generally come from within the organization rather than from external legal or regulatory bodies. But serious misbehavior is also intentional, and thus it is unlikely to be caused purely by naivety or lack of knowledge. Responses must therefore go beyond instructional training, feedback, or other learning-oriented approaches used to correct behaviors that led to unintentional mistakes.

Organizations typically have policies that provide managers with some general guidance about how to act when misbehavior occurs. But these policies often leave much of the decision-making up to individual managers. For example, the Bank of England's staff handbook specifies that if an employee misbehaves, that employee's supervisor or manager will give a first written warning. A failure to fix the misbehavior may lead to a final written warning. After that,

if [the employee's] conduct . . . remain[s] unsatisfactory in any respect . . . it may be decided that further disciplinary action is necessary. This may take the form of dismissal with or without notice. . . . demotion and/or reduction in pay, transfer

Table 1. Types of wrongdoing

Many types of bad behavior can occur in organizations. We focus on the middle category, serious misbehavior, as managers have the most discretion over this form of bad behavior, yet organizational policies often provide insufficient guidance for how managers should respond.

Category of behavior	Subcategory	Example	Typical response
Unintentional mistakes ^A		<ul style="list-style-type: none"> Errors Negligence Unintentional harmful missteps 	<p>Managerial responses focused on improvement, which may include</p> <ul style="list-style-type: none"> performance improvement plans training feedback
Serious misbehavior ^{A,B}	Organizational (the misbehavior primarily harms the functioning of the organization)	<ul style="list-style-type: none"> Employee theft, pilfering^{C,D} Misuse of company resources Vandalism^E Insubordination Substance use^F Sabotage (as in intentionally damaging equipment) Absenteeism^D Misrepresenting work hours 	<p>Managerial responses, possibly guided by organizational policies, that may include</p> <ul style="list-style-type: none"> imposing a negative consequence on the employee attempted offender reintegration
	Group (the misbehavior harms an individual and impairs group or organizational functioning)	<ul style="list-style-type: none"> Discrimination (racism, sexism) Customer harassment^G Social undermining (malicious gossip, rumors, and the like) 	<p>Managerial responses, possibly guided by organizational policies, that may include</p> <ul style="list-style-type: none"> imposing negative consequences on the employee attempted restitution for victims and offender reintegration
	Interpersonal ^H (the misbehavior harms an individual)	<ul style="list-style-type: none"> Harassment of a coworker Bullying Incivility Abusive behavior 	<p>Managerial responses, possibly guided by organizational policies, that may include</p> <ul style="list-style-type: none"> imposing negative consequence on the employee attempted restitution for victims
Criminal misconduct		<ul style="list-style-type: none"> Fraud Embezzlement Extortion, blackmail White-collar crime^I Bribery Assault Ponzi schemes Insider trading False reporting on legal documents 	<p>Termination, responses by external parties (such as the criminal justice system or regulatory agencies), or both</p>

A. Vardi, Y., & Wiener, Y. (1996). Misbehavior in organizations: A motivational framework. *Organization Science*, 7(2), 151–165. <https://doi.org/10.1287/orsc.7.2.151>

B. Ackroyd, S. (2007). Organizational misbehavior. In G. Ritzer (Ed.), *The Blackwell encyclopedia of sociology* (Vol. 18, p. 50). John Wiley & Sons.

C. Greenberg, J. (1990). Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts. *Journal of Applied Psychology*, 75(5), 561–568. <https://doi.org/10.1037/0021-9010.75.5.561>

D. Everton, W. J., Jolton, J. A., & Mastrangelo, P. M. (2007). Be nice and fair or else: Understanding reasons for employees' deviant behaviors. *Journal of Management Development*, 26(2), 117–131. <https://doi.org/10.1108/02621710710726035>

E. DeMore, S. W., Fisher, J. D., & Baron, R. M. (1988). The equity-control model as a predictor of vandalism among college students. *Journal of Applied Social Psychology*, 18(1), 80–91. <https://doi.org/10.1111/j.1559-1816.1988.tb00007.x>

F. Trice, H. M., & Sonnenstuhl, W. J. (1988). Drinking behavior and risk factors related to the work place: Implications for research and prevention. *The Journal of Applied Behavioral Science*, 24(4), 327–346. <https://doi.org/10.1177/002188638802400403>

G. Singer, N. (2020, June 15). Ex-eBay workers sent critics live roaches and a mask of a bloody pig face, U.S. says. *The New York Times*. <https://www.nytimes.com/2020/06/15/technology/ebay-cyberstalking-with-cockroaches-and-bloody-pig-face.html>

H. Robinson, S. L., Wang, W., & Kiewitz, C. (2014). Coworkers behaving badly: The impact of coworker deviant behavior upon individual employees. *Annual Review of Organizational Psychology and Organizational Behavior*, 1(1), 123–143. <https://doi.org/10.1146/annurev-orgpsych-031413-091225>

I. Braithwaite, J. (1985). White collar crime. *Annual Review of Sociology*, 11, 1–25. <https://doi.org/10.1146/annurev.so.11.080185.000245>

“managers are not always able or willing to fire employees”

to other duties or disciplinary suspension without pay.¹⁴

Many other organizations have similarly worded policies, which simultaneously give managers discretion but very little guidance regarding how they should respond.

In addition, most organizational policies fail to acknowledge that managers must consider multiple perspectives beyond the offender’s point of view in selecting their course of action. Employee misbehavior is frequently an open secret. It often occurs in public, is witnessed by others, or becomes known through gossip and rumors or managerial responses. Such behavior can therefore have implications for victims, other employees, and even human resources (HR) departments. As a result, these varied individuals often want to see justice served or restored in their workplace—meaning they are aware that wrongdoing has occurred in their organization and they want steps taken to right the situation. (Scholars refer broadly to this desire as a need for *restoring justice* after a transgression.) Yet policies typically do not tell managers what to do vis-à-vis managing others’ impressions and desires for justice.

In this article, we aim to provide guidance about how managers should respond to serious employee misbehavior, particularly given the multiple viewpoints involved. We walk through the decisionmaking process in choosing and

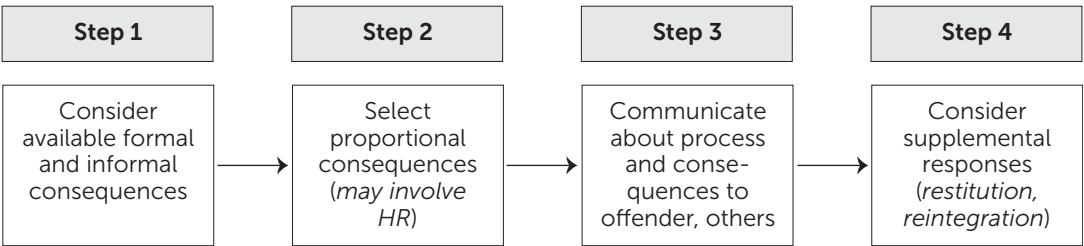
implementing consequences for serious misbehavior. Proceeding in chronological order, per Figure 1, we offer recommendations at each stage: Identifying what actions are available, deciding how to choose among them, and communicating about the chosen response. Along the way, we highlight nine recommendations that are based on our review of existing research. We close with a discussion of supplemental actions that managers can take to further ensure justice is served in the workplace and that victims and wrongdoers find ways to move forward.

Why Delivering Justice Should Be a Priority in Responding to Employee Misbehavior

Imagine you are a manager who discovers that one of your employees has been racially discriminating against colleagues in your organization. This serious misbehavior harms these colleagues and can influence other employees’ behavior in negative ways. Research suggests, for instance, that discrimination based on stereotypes about entire groups of people can be contagious.¹⁵ Furthermore, such behavior can lead to a hostile work environment and send a message that undermines or contradicts the organization’s stated values. In short, this misbehavior causes damage at both the interpersonal and the organizational levels.

Termination may seem to be an obvious response. But managers are not always able or willing to fire employees, even as a response to serious misbehavior, as when the offending employee is protected by tenure or by laws that make it difficult to terminate employees.^{16,17} Even

Figure 1. Chronological decisionmaking process for responding to misbehavior



when managers are able to fire offenders, they often find it deeply uncomfortable to administer this form of punishment.¹⁸ Thus, termination is a relatively rare outcome. Less than 1% of federal employees are fired for misbehavior, for example.^{19,20}

Our first recommendation for managers supervising someone who commits serious misbehavior relates to these cases:

***Recommendation 1.** When employees engage in misbehavior and termination is not an option, managers should find ways to impose other consequences on the offender.*

There are many important reasons to pursue this course of action. For one, it can deter future wrongdoing and help individual victims feel heard and protected. In addition, it sends a powerful message to others in the organization who may be watching events unfold to determine whether theirs is a just workplace.

Why else might it be beneficial for managers to impose consequences? When other employees become aware that misbehavior has occurred, they often desire both retribution²¹ and deterrence to ensure the misbehavior does not happen again.²² Employees therefore look to managers to impose consequences.²³ If employees do not see the offender being held accountable, they are likely to experience a strong sense of injustice, which reduces their motivation,²⁴ productivity,²⁵ and trust in the organization.²⁶

A lack of consequences increases the likelihood that employees will engage in counterproductive work behaviors like destroying equipment, spreading rumors, and stealing.²⁵ Employees may retaliate against managers for failing to denounce misbehavior,²⁷ or they may even attempt to take justice into their own hands and become rogue workplace vigilantes.²⁸ Alternatively, employees may imitate the offender's bad behavior because they believe they can get away with it.²⁹ In short, when managers do not punish misbehavior, this leniency might further damage the organization over time.

The perspectives of other employees also come into play when considering nonpunitive responses to serious misbehavior. Managers may be tempted to outsource punishment to HR departments or rely on developmental approaches, such as education or training aimed at improving an offender's behavior. However, research reveals that third parties prefer punishment to rehabilitation or restoration in the aftermath of serious misbehavior.²² Thus, even if managers impose developmental consequences (like remedial training) or defer to HR departments for punitive actions, other employees will likely believe that the manager responded inadequately to the misbehavior if punitive consequences are not also imposed.

What Formal & Informal Consequences Could a Manager Impose?

A manager needs to assess several aspects of a situation before deciding how to respond to serious misbehavior. Because people have a strong desire for justice and developmental options are unlikely to satisfy this desire, managers should start by considering options that incorporate some form of punishment. Although managers might assume that punitive responses must be severe or involve financial penalties, such as reduced pay, research shows that other types of punishments may be just as effective in deterring future violations and restoring perceptions of justice in an organization.³⁰ In Table 2, we provide a list of punitive consequences that managers could impose.

A number of these consequences can be classified as *formal sanctions*, meaning they make use of existing organizational channels, like HR or payroll processes. These approaches can include demotion,^{31,32} formal written warnings,^{33,34} temporary suspensions,³³ or the revocation of workplace perks like travel or the use of company vehicles. Managers can vary the harshness of these sanctions. For example, a temporary two-day suspension is less severe than an indefinite suspension without pay.

Selecting formal consequences can—and sometimes must—be done in conjunction with

Table 2. Types of consequences available to managers

Consequence type	Example	Description
Formal	Demotion ^{A,B}	Reducing the rank or role of an employee, permanently or temporarily
	Formal written warning or point deduction ^{C,D}	Submitting a formal report to the human resources department or formally deducting points from the employee's points-based evaluation record
	Temporary suspension ^C	Removing the offender from the environment for an extended period of time, either paid or unpaid
	Revocation of perks	Disallowing things from the job that the offender enjoys, such as travel
Informal	Informal warning ^C	Reprimanding the offender without documenting the communication
	Assigning undesirable tasks ^{E,F}	Giving the offender disliked tasks, reassigning desirable or high-status work, or both
	Status reduction ^{G,H,I}	Reducing the influence or esteem of the offender within the group (for example, moving the offender to a less visible workspace, like a basement office)
	Publicly denouncing the behavior in a group communication	Communicating to others that the misbehavior occurred, is unacceptable, and is being responded to, with or without identifying the offender
Informal, to be used with caution (once initiated, these may be difficult for a manager to control)	Embarrassment ^{J,K,L}	Making the offender feel scrutinized by others (for example, allowing gossip to spread)
	Ostracism ^M	Socially excluding the offender (for example, not inviting the offender to important meetings or removing the individual from key committees)
	Public shaming ^{N,O}	Informing others about the offender's transgression while making public statements condemning the transgression, the offender, or both

- A. Carson, P. P., & Carson, K. D. (2007). Demystifying demotion: A look at the psychological and economic consequences on the demotee. *Business Horizons*, 50(6), 455–466. <https://doi.org/10.1016/j.bushor.2007.05.003>
- B. Giacalone, R. A., & Payne, S. L. (1995). Evaluation of employee rule violations: The impact of impression management effects in historical context. *Journal of Business Ethics*, 14(6), 477–487. <https://doi.org/10.1007/BF00872088>
- C. Beyer, J. M., & Trice, H. M. (1984). A field study of the use and perceived effects of discipline in controlling work performance. *Academy of Management Journal*, 27(4), 743–764. <https://doi.org/10.5465/255876>
- D. Herath, T., & Rao, H. R. (2009). Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. *Decision Support Systems*, 47(2), 154–165. <https://doi.org/10.1016/j.dss.2009.02.005>
- E. Hoffer, J. A., & Straub, D. W., Jr. (1989). The 9 to 5 underground: Are you policing computer crimes? *MIT Sloan Management Review*, 30(4), 35–43.
- F. Straub, D. W., & Welke, R. J. (1998). Coping with systems risk: Security planning models for management decision making. *MIS Quarterly*, 22(4), 441–469.
- G. Gephart, R. P. (1978). Status degradation and organizational succession: An ethnomethodological approach. *Administrative Science Quarterly*, 23(4), 553–581.
- H. Pettit, N. C., Yong, K., & Spataro, S. E. (2010). Holding your place: Reactions to the prospect of status gains and losses. *Journal of Experimental Social Psychology*, 46(2), 396–401. <https://doi.org/10.1016/j.jesp.2009.12.007>
- I. van Prooijen, J.-W., & Lam, J. (2007). Retributive justice and social categorizations: The perceived fairness of punishment depends on intergroup status. *European Journal of Social Psychology*, 37(6), 1244–1255. <https://doi.org/10.1002/ejsp.421>
- J. Feinberg, M., Willer, R., Stellar, J., & Keltner, D. (2012). The virtues of gossip: Reputational information sharing as prosocial behavior. *Journal of Personality and Social Psychology*, 102(5), 1015–1030. <https://doi.org/10.1037/a0026650>
- K. Grasmick, H. G., Bursik, R. J., Jr., & Kinsey, K. A. (1991). Shame and embarrassment as deterrents to noncompliance with the law: The case of an antilittering campaign. *Environment and Behavior*, 23(2), 233–251. <https://doi.org/10.1177/0013916591232006>
- L. Okimoto, T. G., & Wenzel, M. (2011). Third-party punishment and symbolic intragroup status. *Journal of Experimental Social Psychology*, 47(4), 709–718. <https://doi.org/10.1016/j.jesp.2011.02.001>
- M. Williams, K. D. (2007). Ostracism: The kiss of social death. *Social and Personality Psychology Compass*, 1(1), 236–247. <https://doi.org/10.1111/j.1751-9004.2007.00004.x>
- N. Andrighetto, G., Brandts, J., Conte, R., Sabater-Mir, J., Solaz, H., & Villatoro, D. (2013). Punish and voice: Punishment enhances cooperation when combined with norm-signalling. *PLOS ONE*, 8(6), Article e64941. <https://doi.org/10.1371/journal.pone.0064941>
- O. Harris, M. A. (2012). Shaming as a technique for information security policy and training adherence. *SAIS 2012 Proceedings*, Article 16. <https://aisel.aisnet.org/sais2012/16>

HR departments. This collaboration may mean that managers do not have complete discretion over how to respond. Moreover, managers and HR personnel might disagree about the best course of action, because HR personnel sometimes have different incentives than managers do. For example, if members of the HR department are concerned about offenders suing the organization, they may prevent managers from imposing proportional consequences on offenders.³⁵ Managers may be tempted to simply accept the HR department's approach—indeed, they might even feel relieved to be absolved of the responsibility of responding. However, if others believe the HR department's actions are insufficient, then managers will have to respond in other ways or communicate about the ways in which they are constrained (see the *How Should Consequences Be Communicated?* section).

Managers may also administer *informal sanctions*, which convey social disapproval but do not involve formal channels. For example, managers who feel they need to take action beyond the HR department's response might choose to give an offender less desirable work assignments.³⁶ Among other informal sanctions that managers can impose are verbal reprimands;³³ reductions in status, prestige, or social standing;^{37,38} ostracization of the offender;³⁹ and even public shaming.^{40,41} Research suggests that informal sanctions can be just as effective as formal sanctions at deterring future transgressions⁴² and improving the workforce's perceptions of justice.

Informal sanctions do not require organizational support or administrative processes; managers (and even peer employees) can impose them when their own status, position, or political capital does not otherwise allow them to pursue formal sanctions. In fact, managers may have to use informal sanctions when formal channels are unavailable. A downside to some of these approaches is that once they are enacted, it is difficult or impossible to stop or constrain them.⁴³ For example, severe public shaming in response to a relatively minor offense (like arriving late to a meeting) could lead to social ridicule and turn the offender into a pariah.

Because informal sanctions often arise organically, managers should find out what informal sanctions are already happening within their organization so that they do not excessively pile on additional consequences. (They may also need to intervene if peers' sanctions have gone too far.) These concerns should not deter managers from using informal sanctions. Rather, they are reminders that managers should also think through how informal sanctions can be constrained. In Table 2, we flag particular informal sanctions that should be used with caution.

Making Decisions About Appropriate Consequences

In addition to determining whether formal or informal sanctions are appropriate, managers need to consider the severity of the misbehavior. How much harm was caused, or how "wrong" is the misbehavior perceived to be? Managers need to evaluate multiple perspectives on these questions to satisfy their employees' sense that the consequences are proportional to the misbehavior (that is, sanctions are neither too mild nor too harsh).

Third parties may feel justice has not been restored when consequences seem too lenient.⁴⁴ And draconian managerial punishments may lead third parties to believe that the offender has not been treated fairly.⁴⁵ Furthermore, offenders who feel unjustly treated may retaliate in some way.⁴⁶ In other words, responses to employee misbehavior must be harsh but fair enough to both restore justice in the eyes of others in the organization and avoid creating offender backlash.

Making the selection of consequences even more challenging is the fact that assessments of wrongdoing, severity, and proportionality are highly subjective. Such judgments are influenced by many factors, such as culture,⁴⁷ knowledge of the situation, individuals' varying perspectives, and the capacity to empathize with others.⁴⁸

This leads to our second recommendation:

***Recommendation 2.** Managers should select consequences that—in the eyes of others—are proportional to the severity of the misbehavior.*

In other words, managers must impose consequences that are considered appropriate by both offenders and others in the organization. To do so, managers should consult the victim or victims (if specific victims are involved), the offender or offenders, and other employees to learn about and understand what they would consider to be an appropriate response. Insights into this approach come from research on procedural justice—that is, the study of whether the procedures used to arrive at decisions and outcomes are perceived as fair. These studies show that decisions made with an emphasis on transparency and in consultation with others who might care about the decision (such as disciplinary boards) facilitate buy-in on the decision.⁴⁹

What to Do When Facing Uncertainty Regarding the Misbehavior

Uncertainty surrounding a purported misbehavior can make it difficult to impose proportional consequences. Uncertainty can take several forms. In some cases, it is hard to determine whether a harmful behavior actually occurred. For example, managers might not be able to verify accusations of inappropriate sexual comments, or they might find that people disagree about whether a lunch invitation was, in fact, a sexual proposition. In such situations, managers should first gather all relevant information from reporting parties and other observers. While collecting these details, managers must remember that victim and third-party perceptions of whether misbehavior occurred are what matter. If the victim and third parties perceive that misbehavior occurred and a managerial response is absent, these individuals are likely to feel that justice has not been restored.

Thus, we offer our third recommendation:

***Recommendation 3.** When managers cannot verify whether harmful behavior*

occurred, they should communicate to the reporting employee (a) that they have heard the allegation, (b) the reasons why they cannot verify whether misbehavior occurred, (c) what they intend to do in response, and (d) how the uncertainty influenced their decision.

This multistep approach is powerful in part because it is transparent—and several studies have demonstrated that employees care about transparency and procedural justice.⁴⁹

Uncertainty can also arise when the standards regarding organizational values, norms, or policies are unclear or if the seriousness of the misbehavior is subject to debate. For instance, stealing office supplies or occasional tardiness may violate rules but, in certain organizations, these offenses may not be classified as serious.

In these scenarios, managers should start by seeking clarification from the HR department about their organization's standards and use their discretion when deciding whether the behavior constitutes serious misbehavior. Managers can also seek feedback from trusted parties, both formally and informally, to better gauge how others perceive the situation. Once again, what matters most is that the approach is transparent and consistent across offenders who commit similar transgressions. To the extent that managers judge serious misbehavior to have occurred, punitive responses should be enacted. However, if there are mitigating circumstances (for example, the offender was coerced into the misbehavior), managers could consider applying less severe penalties.

Finally, managers may experience uncertainty about whether the offender intended to seriously misbehave. Regardless of the offender's underlying intentions, if serious misbehavior occurred, punishment should be enacted to uphold the organization's values⁵⁰ and because third parties are likely to perceive a lack of punishment to be unfair. Our next recommendation comes into play in these situations:

***Recommendation 4.** When misbehavior could be partly unintentional, managers*

should consider supplementing their punitive response with a developmental response.

For example, when British soccer player Jamie Vardy used a racist slur against a Japanese man in a casino, Vardy said that he was unaware that the term could be offensive.⁵¹ The athletic club of which Vardy was a part needed to convey that racism violates their values and would not be tolerated. But given Vardy's professed ignorance, the club leadership supplemented their punitive response with a developmental one by both fining him and requiring diversity training.⁵¹ This approach meets the expectations of third parties, who want to see punitive action when serious misbehavior occurs, and also accommodates the perspective of offenders, who typically feel that because their misbehavior was unintentional, they should be treated leniently or receive rehabilitative treatment such as education or training. Furthermore, when coupled with a developmental approach, punishment may feel more palatable to managers.

How Should Consequences Be Communicated?

Before managers communicate their course of action, they should first consider how best to approach the offender. A punitive consequence can make offenders feel marginalized and unvalued,¹⁰ so offenders may psychologically withdraw from the organization afterward.⁵² Managers, meanwhile, may need offenders to remain involved and productive members of the organization. In these cases, we recommend the following:

***Recommendation 5.** If managers wish to keep offenders engaged at the organization, they should communicate consequences to offenders using language that makes offenders feel like they still belong at the organization.*

Organizational behavior research has found that managers can simultaneously punish offenders while also using *reintegrative language*^{10,53} to frame the manager's decisions in ways that affirm that offenders are still valued and

"employees might assume that the offender got away with the misbehavior"

can recover from the incident. For example, a manager might tell an offender that "your actions are not who you are. Do not let them define you. People mess up, and it's not the mess-ups that define them, it's how they come back from them."¹⁰

Not only is it important to communicate consequences to offenders, but it is equally important that managers inform other employees that the offender has been held accountable.⁵⁴ Without such communication, employees might assume that the offender got away with the misbehavior, leading them to perceive the workplace as being unjust,⁵⁴ even though the manager has, in fact, administered discipline. We therefore offer our sixth recommendation:

***Recommendation 6.** Managers should communicate to others in the organization that they have taken actions to hold the offender responsible for the misbehavior.*

In some circumstances, managers cannot speak directly about an offender—as when legal concerns or HR policies bar this option—but managers can still share their responses to the misbehavior with employees by making a general statement about the type of misbehavior and its consequences without mentioning an individual situation or naming the offender. Studies have found that even indirectly communicating that an offender was held accountable—without specifying the exact actions taken—may help employees feel that justice has been restored.⁵⁰

When updating observers in this way, managers should ensure that the following two elements are a part of their message:

***Recommendation 7.** Managers should clearly convey that the consequences*

were a direct result of the misbehavior (that is, they were contingent on the behavior), and managers should supply the reason for the consequences (the rationale).

In the aftermath of bad behavior, people spontaneously and automatically attempt to understand why the events occurred,^{55,56} in part to anticipate a manager's choices and the likelihood of these transgressions occurring in the future. For this reason, managers who directly link misbehavior with negative consequences (through *contingent communication*) can help other employees understand the consequences and their context.⁵⁷

One example of contingent communication is a message from Chicago's National Hockey League team, the Blackhawks. When the league suspended the team's assistant coach, Marc Crawford, for verbally abusing players, the organization issued a statement:

[We, along with] independent legal counsel, conducted a thorough review of assistant coach Marc Crawford in response to allegations of misconduct in previous coaching positions. . . .

We do not condone his previous behavior. Through our review, we confirmed that Marc proactively sought professional counseling to work to improve and become a better communicator, person and coach. . . .

We have determined that Marc will remain suspended from team activities until January 2, 2020, at which time he will resume his assistant coaching duties, subject to his continued compliance with his contractual obligations and team expectations. In addition, he will continue with his counseling moving forward.⁵⁸

This statement communicates that Crawford was held accountable and that his suspension was directly linked to his alleged misconduct. When consequences are understood to be contingent on a particular instance of misbehavior, observers infer that misbehavior will lead

to consequences. This cause–effect pairing helps observers feel that they understand the rules of the organization, which fosters perceptions of a just work environment.⁵⁹ Moreover, employees who feel that consequences in the workplace are contingent on misbehavior perceive their supervisors to be more effective.⁵⁷ Other research indicates that workplaces with contingent discipline are often more functional than workplaces with noncontingent discipline.⁶⁰

In addition to articulating contingency, managers should offer a rationale for choosing those consequences that clarifies the intended purpose or goal of an intervention. Take the case of a Southwest pilot caught making sexist remarks about flight attendants. Southwest temporarily suspended the pilot and sent him to diversity training. A vice president from Southwest announced these consequences, saying that the goal of the suspension and training was “to reinforce the company's expectation that [the pilot] show respect and treat all with dignity.”⁶¹

Absent a communicated rationale, employees may view the process of administering consequences to be unfair, which could be problematic: Perceptions of unjust processes harm organizational functioning even more than do perceptions of unjust outcomes.⁴⁹ Failing to explain the rationale behind a manager's response to misbehavior may lead observers to infer that the consequences were being administered for a different reason. This misunderstanding may then lead them to view the consequences as mismatched to the misbehavior and therefore ineffective. For that matter, if observers misunderstand the manager's intentions, they may conclude the intended goals were not achieved. A manager may choose to move a misbehaving employee to a new role in which it is impossible for the offender to reoffend, for example. But if other employees believe that the relocation was the manager's effort to punish the offender in a way that would make the transgressor suffer for their crime, they may see this response as too soft, thus failing to understand the true rationale behind the manager's actions.

What Can Managers Do to Provide Restitution to Victims of Serious Misbehavior?

When a person's misbehavior harms another individual, the victim of this interpersonal misbehavior may feel that their status, belonging, autonomy, dignity, and respect have been compromised.^{5,62} As a result, victims (and other employees) tend to perceive the social dynamics of an organization as being unbalanced after interpersonal misbehavior.^{63,64} For example, they often feel the victim is owed something from the offender or the organization broadly. Offender-focused consequences might be one way to reestablish justice perceptions, but managers should also consider victim-focused responses.⁶⁵ Hence, beyond attempting to have the offender experience proportional consequences, managers may also need to think about how best to accomplish *victim restitution*—that is, how best to provide compensation to or otherwise help to assuage the distress of the wronged party. This leads to our next recommendation:

Recommendation 8. *Managers should consider actions that support victims, such as compensatory justice and apologies.*

Victims care greatly about having their status restored and needs met in the wake of interpersonal misbehavior.⁶⁶ Moreover, substantial research shows how important it is to both adopt procedures that ensure respect for victims (procedural justice)⁴⁹ and ensure that victims feel respected and heard (interactional justice).⁶⁷ The enactment of these responses, which can restore the victim's understanding that they work within a just organization, may reduce the likelihood that the victims of wrongdoing retaliate against the organization.⁶⁸ More generally, managers must create a psychologically safe environment in which people can bring problems to the attention of management without fear.⁶⁹ The absence of such a climate, by contrast, works against managers' efforts to restore perceptions of justice and can lead to the proliferation of additional misbehavior.^{70,71} In sum, after interpersonal misbehavior occurs, managers should not only impose

consequences on the offender but also publicly and visibly take actions that address victims' needs and rights.

The rights of victimized employees are often enumerated in employee handbooks—they may even have access to ombudsmen or the right to a union representative to aid them—but some formal organizational policies actually limit victims' rights in the aftermath of misbehavior. Organizations are often fearful of legal action by victims—which could harm the reputation of the organization or the offenders—and they therefore put obstacles in place to deter lawsuits or protect the status of either the organization or offenders.⁷² For example, sexual harassment cases are often settled under a nondisclosure agreement (NDA) in an attempt to protect the reputation of the offender.⁷³ However, such practices do little to aid victims.

Putting aside concerns about the fairness and morality of NDAs in this context, it is important to consider what might actually help victims feel psychologically protected and restored and to avoid any attempts to offer mere appeasement. One option is to demonstrate *active compassion* for the victim, meaning to take steps that offer comfort or otherwise address the victim's pain.⁷⁴ Managers can do so by giving victims the time and resources to pursue their own path to restoring justice, providing forums for victims to air their grievances and voice their experiences,⁷⁵ encouraging offenders to apologize to victims,^{76,77} and providing financial or other forms of compensation.^{78,79}

The exact actions managers take on behalf of victims may depend in part on the nature of the misbehavior. Victims may want face-to-face apologies if they were treated unfairly—but not if they suffered sexual harassment or other traumatizing offenses.⁸ In the former case, managers could provide forums for offenders to apologize. In the latter, managers could move the offender out of the victim's work environment and provide the victim with compensatory resources, such as a private forum in which to talk about the harassment experienced, time off from work, social support, or sponsorship during a job search.



Annual loss from
employee theft in the
United States is
\$50 billion

<1%

Federal employees
fired for misbehavior

“victims themselves should be involved in the justice process”

Managers should also consider involving people at multiple levels of the organization in the response. Although managers who directly supervise the victim and offender—and therefore know both parties well—may be more likely than senior leaders to gain offender cooperation,⁸⁰ the involvement of senior leaders can foster perceptions of justice in other ways. If senior leaders are involved in selecting and communicating responses to wrongdoing, victims may feel that the organization is taking their perspective seriously. The involvement of high-status leaders gives employees the impression that justice—and the specific incident at hand—is important to the organization.^{81,82} Most critically, victims themselves should be involved in the justice process: As noted previously, managers should ask the wronged individuals for their input on the situation, particularly about what kind of response or punishment is warranted. In addition, managers should make victims aware that their actions are intended to restore justice and empower the victim within the organization. Otherwise, victims or third parties may perceive such actions to be attempts at victim appeasement rather than a sincere effort to be supportive.

If managers handle these responses well, they may even accrue some personal benefits. Managers who both compensate victims and discipline offenders are held in high esteem by observers. They are more likely to garner support for election to a specific office or promotion within an organization than are managers who only discipline offenders.⁶⁵

How Can Managers Reintegrate Offenders?

Offenders who remain at an organization pose many challenges that managers should not ignore. As described earlier, wrongdoers may disengage from the organization after being punished. Alternatively, other employees may not accept offenders who remain, which

could lead to perceptions of injustice, dysfunctional interactions, and loss of productivity.⁸³ Because organizations depend on employee engagement and cooperation, managers may want to support *offender reintegration*,¹⁰ that is, the repair of relationships shattered by the transgression so that the offender can again become an accepted part of the organizational community.⁸⁴

For example, managers can help people who have committed serious misbehavior reestablish trust with others by facilitating their efforts to actively make amends^{53,84} that go above and beyond the consequences imposed by the manager, whether in expressing sincere concern for third parties or by publicly reaffirming the organization's values.⁴⁸ If third parties respond in supportive ways to the offenders' repair attempts, offenders, in turn, are likely to feel reaccepted by organizational members.^{10,48,84}

To foster reintegration, managers can bring offenders and other organizational members together through *restorative justice conferencing*,^{48,75} in which offenders offer amends and third parties reaffirm support. This discussion can also serve as a space for third parties to voice their grievances, which can be therapeutic and facilitate relationship repair.⁸⁵ Managers can also support offender reintegration by making public statements describing the restorative justice conferencing process,^{48,75} affirming the offender's continued role in the organization, or both.

These steps can help offenders feel reaccepted and can clarify to third parties how they should interact with offenders. For example, Louisiana State University (LSU) suspended basketball coach Will Wade for violating the National Collegiate Athletic Association's (NCAA's) policies for recruiting student-athletes and initially declining to meet with administrators. When there was no evidence of misconduct and Wade was later reinstated, the LSU athletic director made a public statement reaffirming his place:

The LSU Athletics Department today agreed to reinstate Will Wade as head coach of the Tiger basketball program.

Coach Wade met Friday with University and NCAA officials. During those meetings, he answered all questions . . . in connection with recently reported allegations of irregularities in college basketball recruiting.

The University regrets that Coach Wade did not choose to fulfill his obligations to LSU when he was first asked to do so. However, the seriousness of the allegations and Coach Wade's prior refusal to refute them could not be ignored without exposing the University and the basketball program to great risk. Protecting LSU and preserving our integrity must always be our first priority.

Coach Wade's explanations and clarifications offered during the meeting, absent actual evidence of misconduct, satisfy his contractual obligation to LSU. Accordingly, I have recommended that Coach Wade's suspension be lifted and that he should be allowed to resume his coaching responsibilities.⁸⁶

Managers can foster reintegration in some cases by giving offenders a chance to publicly address their misbehavior. This strategy is not necessarily appropriate in cases where there are clear victims, but it can be effective for misbehavior that harmed an organization rather than individuals. In the case of LSU, Coach Wade issued a public statement:

I am humbled and grateful to be back at LSU. I would like to express my appreciation to [the] President . . . and Athletic Director . . . for my reinstatement, and I sincerely apologize to the university and our fans for the disruption to the University and the program.

I regret the circumstances that prevented me from meeting with the University sooner. I wish I could have addressed these issues when the University first requested a meeting, and I'm grateful they gave me the opportunity to do so last week.

I completely understand that without my denying or explaining the media reports accusing me of wrongdoing LSU

was left with no choice but to suspend me until I was willing and able to meet with them. Any other course of action would have put the program and the University at risk.

I look forward to re-joining the team right away. I intend to sit down with my student-athletes and co-workers to explain what has happened during the last 30 days and how I intend for us all to move forward.⁸⁷

This example illustrates how offenders can publicly take responsibility, express remorse, and describe their plans for future action and improvement, which can facilitate reintegration and reacceptance. Even if HR departments or legal concerns prevent managers or wrongdoers from making public statements, managers may be able to articulate such information privately to their work groups or orchestrate private opportunities for offenders to communicate with other organizational members.

We summarize our advice for offender reintegration with a final recommendation:

***Recommendation 9:** Managers should carefully consider whether to attempt to reintegrate offenders; if they decide to do so, they can accomplish reintegration by (a) facilitating interactions between offenders and other organizational members, (b) making public or private statements to others that affirm the offenders' place (and future) at the organization, (c) offering offenders a chance to voice apologies to others in the organization, or (d) doing some combination of these.*

We must caution that offender reintegration is not always appropriate. If employees believe that an offender should have been terminated or received a harsher consequence than was actually administered, seeing the offender being actively reintegrated into an organization may increase employees' perceptions of injustice.⁴⁴ Thus, offender reintegration should only be pursued if managers believe that other employees have been satisfied by the negative

consequences administered. In addition, reintegration may be more suitable in the case of organizational misbehavior, when no identifiable individual victims exist.⁶² Third parties are generally highly sympathetic to victims;^{88,89} thus, if employees believe that offender reintegration is taking precedence over victim restitution, they are likely to believe that justice is not being restored.

Conclusion

In the aftermath of employee misbehavior in organizations, when an offender remains employed, managers must find ways to hold the offender accountable and foster perceptions of justice among employees. By selecting appropriate consequences, communicating those consequences to everyone involved, and taking actions to support both victim restitution and offender reintegration, managers can help restore and enhance perceptions of justice in the organization. Disciplining offending employees and addressing the concerns of victims and other observers may not be easy, but these steps are critical for moving organizations successfully through incidents of bad behavior.

author affiliation

Frey: University of Southern California. Adams & Detert: University of Virginia. Bruno: James Madison University. Corresponding author's email: elfrey@marshall.usc.edu.

references

1. Memmott, M. (2010, February 3). *Should Australian 'porn broker' be disciplined?* NPR. https://www.npr.org/sections/thetwo-way/2010/02/porn_australian_banker_surfing.html
2. Pofeldt, E. (2017, September 12). *This crime in the workplace is costing US businesses \$50 billion a year*. CNBC. <https://www.cnbc.com/2017/09/12/workplace-crime-costs-us-businesses-50-billion-a-year.html>
3. Cattaneo, A. A., & Boldrini, E. (2017). You learn by your mistakes: Effective training strategies based on the analysis of video-recorded worked-out examples. *Vocations and Learning*, 10(1), 1–26. <https://doi.org/10.1007/s12186-016-9157-4>
4. Vardi, Y., & Wiener, Y. (1996). Misbehavior in organizations: A motivational framework. *Organization Science*, 7(2), 151–165. <https://doi.org/10.1287/orsc.7.2.151>
5. Tepper, B. J., & Henle, C. A. (2011). A case for recognizing distinctions among constructs that capture interpersonal mistreatment in work organizations. *Journal of Organizational Behavior*, 32(3), 487–498. <https://doi.org/10.1002/job.688>
6. Bennett, R. J., & Robinson, S. L. (2000). Development of a measure of workplace deviance. *Journal of Applied Psychology*, 85(3), 349–360. <https://doi.org/10.1037/0021-9010.85.3.349>
7. Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology*, 92(2), 410–424. <https://doi.org/10.1037/0021-9010.92.2.410>
8. Lim, S., & Cortina, L. M. (2005). Interpersonal mistreatment in the workplace: The interface and impact of general incivility and sexual harassment. *Journal of Applied Psychology*, 90(3), 483–496. <https://doi.org/10.1037/0021-9010.90.3.483>
9. Fox, S., Spector, P. E., & Miles, D. (2001). Counterproductive work behavior (CWB) in response to job stressors and organizational justice: Some mediator and moderator tests for autonomy and emotions. *Journal of Vocational Behavior*, 59(3), 291–309. <https://doi.org/10.1006/jvbe.2001.1803>
10. Frey, E. (2022). “I go here...but I don't necessarily belong”: The process of transgressor reintegration in organizations. *Academy of Management Journal*, 65(1), 119–157. <https://doi.org/10.5465/amj.2019.0251>
11. Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (2004). Removing the shadow of suspicion: The effects of apology versus denial for repairing competence- versus integrity-based trust violations. *Journal of Applied Psychology*, 89(1), 104–118. <https://doi.org/10.1037/0021-9010.89.1.104>
12. Aquino, K., Tripp, T. M., & Bies, R. J. (2001). How employees respond to personal offense: The effects of blame attribution, victim status, and offender status on revenge and reconciliation in the workplace. *Journal of Applied Psychology*, 86(1), 52–59. <https://doi.org/10.1037/0021-9010.86.1.52>
13. Treviño, L. K. (2005). The role of leaders in influencing unethical behavior in the workplace. In R. E. Kidwell & C. L. Martin (Eds.), *Managing organizational deviance* (pp. 69–87). SAGE Publications. <https://doi.org/10.4135/9781452231105.n3>
14. Bank of England. (2019). *Bank of England staff handbook*. <https://www.bankofengland.co.uk/-/media/boe/files/about/human%20resources/staff-handbook>
15. Willard, G., Isaac, K.-J., & Carney, D. R. (2015). Some evidence for the nonverbal contagion of racial bias. *Organizational Behavior and Human Decision Processes*, 128, 96–107. <https://doi.org/10.1016/j.obhdp.2015.04.002>
16. Office of Personnel Management. (1999). *Poor performers in government: A quest for a true story*. U.S. Office of Merit Systems Oversight and Effectiveness.
17. Liff, S. (2009). *The complete guide to hiring and firing government employees*. AMACOM.
18. Butterfield, K. D., Treviño, L. K., & Ball, G. A. (1996). Punishment from the manager's perspective: A grounded investigation and inductive model. *Academy of Management Journal*, 39(6), 1479–1512. <https://doi.org/10.5465/257066>
19. Yoder, E. (2017, October 18). Hardly any federal employees are fired for poor performance. That could be a good sign, report says. *The Washington Post*. <https://www.washingtonpost.com/news/powerpost/wp/2017/10/18/hardly-any-federal-employees-are-fired-for-poor-performance-that-could-be-a-good-sign-report-says/>
20. Meyer, A. (2016, August 11). Workers in private sector are 3 times more likely to get fired than gov't workers. *The Washington Free Beacon*. <https://freebeacon.com/issues/workers-private-sector-3-times-likely-get-fired-govt-workers/>
21. Crockett, M. J., Özdemir, Y., & Fehr, E. (2014). The value of vengeance and the demand for deterrence. *Journal of Experimental Psychology: General*, 143(6), 2279–2286. <https://doi.org/10.1037/xge0000018>
22. Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology*, 83(2), 284–299. <https://doi.org/10.1037/0022-3514.83.2.284>
23. Gross, J., Méder, Z. Z., Okamoto-Barth, S., & Riedl, A. (2016). Building the Leviathan—Voluntary centralisation of punishment power sustains cooperation in humans. *Nature Scientific Reports*, 6(1), Article 20767. <https://doi.org/10.1038/srep20767>
24. Colquitt, J. A., & Chertkoff, J. M. (2002). Explaining injustice: The interactive effect of explanation and outcome on fairness perceptions and task motivation. *Journal of Management*, 28(5), 591–610. [https://doi.org/10.1016/S0149-2063\(02\)00157-5](https://doi.org/10.1016/S0149-2063(02)00157-5)
25. Cohen-Charash, Y., & Spector, P. E. (2001). The role of justice in organizations: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 86(2), 278–321. <https://doi.org/10.1006/obhd.2001.2958>
26. DeConinck, J. B. (2010). The effect of organizational justice, perceived organizational support, and perceived supervisor support on marketing employees' level of trust. *Journal of Business Research*, 63(12), 1349–1355. <https://doi.org/10.1016/j.jbusres.2010.01.003>
27. Martin, J. W., Jordan, J. J., Rand, D. G., & Cushman, F. (2019). When do we punish people who don't? *Cognition*, 193, 1–13. <https://doi.org/10.1016/j.cognition.2019.104040>
28. DeCelles, K.A., & Aquino, K. (2020). Dark knights: When and why an employee becomes a workplace vigilante. *Academy of Management Review*, 45(3), 528–548. <https://doi.org/10.5465/amr.2017.0300>
29. Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological Science*, 20(3), 393–398. <https://doi.org/10.1111/j.1467-9280.2009.02306.x>

30. Masclet, D., Noussair, C., Tucker, S., & Villeval, M. C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review*, 93(1), 366–380. <https://doi.org/10.1257/000282803321455359>
31. Carson, P. P., & Carson, K. D. (2007). Demystifying demotion: A look at the psychological and economic consequences on the demotee. *Business Horizons*, 50(6), 455–466. <https://doi.org/10.1016/j.bushor.2007.05.003>
32. Giacalone, R. A., & Payne, S. L. (1995). Evaluation of employee rule violations: The impact of impression management effects in historical context. *Journal of Business Ethics*, 14(6), 477–487. <https://doi.org/10.1007/BF00872088>
33. Beyer, J. M., & Trice, H. M. (1984). A field study of the use and perceived effects of discipline in controlling work performance. *Academy of Management Journal*, 27(4), 743–764. <https://doi.org/10.5465/255876>
34. Herath, T., & Rao, H. R. (2009). Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. *Decision Support Systems*, 47(2), 154–165. <https://doi.org/10.1016/j.dss.2009.02.005>
35. Weaver, G. R., & Treviño, L. K. (2001). The role of human resources in ethics/compliance management: A fairness perspective. *Human Resource Management Review*, 11(1–2), 113–134. [https://doi.org/10.1016/S1053-4822\(00\)00043-7](https://doi.org/10.1016/S1053-4822(00)00043-7)
36. Hoffer, J. A., & Straub, D. W., Jr. (1989). The 9 to 5 underground: Are you policing computer crimes? *MIT Sloan Management Review*, 30(4), 35–43.
37. Gephart, R. P. (1978). Status degradation and organizational succession: An ethnomethodological approach. *Administrative Science Quarterly*, 23(4), 553–581.
38. Pettit, N. C., Yong, K., & Spataro, S. E. (2010). Holding your place: Reactions to the prospect of status gains and losses. *Journal of Experimental Social Psychology*, 46(2), 396–401. <https://doi.org/10.1016/j.jesp.2009.12.007>
39. Williams, K. D. (2007). Ostracism: The kiss of social death. *Social and Personality Psychology Compass*, 1(1), 236–247. <https://doi.org/10.1111/j.1751-9004.2007.00004.x>
40. Andrighetto, G., Brandts, J., Conte, R., Sabater-Mir, J., Solaz, H., & Villatoro, D. (2013). Punish and voice: Punishment enhances cooperation when combined with norm-signalling. *PLOS ONE*, 8(6), Article e64941. <https://doi.org/10.1371/journal.pone.0064941>
41. Harris, M. A. (2012). Shaming as a technique for information security policy and training adherence. *SAIS 2012 Proceedings*, Article 16. <https://aisel.aisnet.org/sais2012/16>
42. Anderson, L. S., Chiricos, T. G., & Waldo, G. P. (1977). Formal and informal sanctions: A comparison of deterrent effects. *Social Problems*, 25(1), 103–114.
43. Wesselmann, E. D., Wirth, J. H., Pryor, J. B., Reeder, G. D., & Williams, K. D. (2013). When do we ostracize? *Social Psychological and Personality Science*, 4(1), 108–115. <https://doi.org/10.1177/1948550612443386>
44. Niehoff, B. P., Paul, R. S., & Bunch, J. F. S. (1998). The social effects of punishment events: The influence of violator past performance record and severity of the punishment on observers' justice perceptions and attitudes. *Journal of Organizational Behavior*, 19(6), 589–602. [https://doi.org/10.1002/\(SICI\)1099-1379\(1998110\)19:6<589::AID-JOB862>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1099-1379(1998110)19:6<589::AID-JOB862>3.0.CO;2-N)
45. Tepper, B. J. (2000). Consequences of abusive supervision. *Academy of Management Journal*, 43(2), 178–190. <https://doi.org/10.5465/1556375>
46. Ball, G., Treviño, L., & Sims, H. (1994). Just and unjust punishment: Influences on subordinate performance and citizenship. *Academy of Management Journal*, 37(2), 299–322. <https://doi.org/10.5465/256831>
47. Tsui, J., & Windsor, C. (2001). Some cross-cultural evidence on ethical reasoning. *Journal of Business Ethics*, 31(2), 143–150. <https://doi.org/10.1023/A:1010727320265>
48. Okimoto, T. G., & Wenzel, M. (2014). Bridging diverging perspectives and repairing damaged relationships in the aftermath of workplace transgressions. *Business Ethics Quarterly*, 24(3), 443–473. <https://doi.org/10.5840/beq201471515>
49. Tyler, T. R., & Lind, E. A. (2002). Procedural justice. In J. Sanders & V. L. Hamington (Eds.), *Handbook of justice research in law* (pp. 65–92). Springer. https://doi.org/10.1007/0-306-47379-8_3
50. Okimoto, T. G., & Wenzel, M. (2009). Punishment as restoration of group and offender values following a transgression: Value consensus through symbolic labelling and offender reform. *European Journal of Social Psychology*, 39(3), 346–367. <https://doi.org/10.1002/ejsp.537>
51. Jamie Vardy 'racist' slur: Striker says he was 'ignorant.' (2020, September 26). BBC. <https://www.bbc.com/news/uk-england-leicestershire-37477703>
52. Huo, Y. J., Binning, K. R., & Molina, L. E. (2010). Testing an integrative model of respect: Implications for social engagement and well-being. *Personality and Social Psychology Bulletin*, 36(2), 200–212. <https://doi.org/10.1177/0146167209356787>
53. Goodstein, J., Butterfield, K., & Neale, N. (2016). Moral repair in the workplace: A qualitative investigation and inductive model. *Journal of Business Ethics*, 138(1), 17–37. <https://doi.org/10.1007/s10551-015-2593-5>
54. Treviño, L. K. (1992). The social effects of punishment in organizations: A justice perspective. *Academy of Management Review*, 17(4), 647–676. <https://doi.org/10.5465/amr.1992.4279054>
55. Kanazawa, S. (1992). Outcome or expectancy? Antecedent of spontaneous causal attribution. *Personality and Social Psychology Bulletin*, 18(6), 659–668. <https://doi.org/10.1177/0146167292186001>
56. Weiner, B. (1985). "Spontaneous" causal thinking. *Psychological Bulletin*, 97(1), 74–84. <https://doi.org/10.1037/0033-2909.97.1.74>
57. Atwater, L. E., Dionne, S. D., Camobreco, J. F., Avolio, B. J., & Lau, A. (1998). Individual attributes and leadership style: Predicting the use of punishment and its effects. *Journal of Organizational Behavior*, 19(6), 559–576. [https://doi.org/10.1002/\(SICI\)1099-1379\(1998110\)19:6<559::AID-JOB866>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1099-1379(1998110)19:6<559::AID-JOB866>3.0.CO;2-E)
58. Roumeliotis, C. (2019, December 16). *Blackhawks, Marc Crawford release statements regarding investigation*. NBC Sports. <https://www.nbcsports.com/chicago/blackhawks/blackhawks-marc-crawford-release-statements-regarding-investigation>
59. Podsakoff, P. M., Bommer, W. H., Podsakoff, N. P., & MacKenzie, S. B. (2006). Relationships between leader reward and punishment attitudes, perceptions, and behaviors: A meta-analytic review of existing and new research. *Organizational Behavior and Human Decision Processes*, 99(2),

- 113–142. <https://doi.org/10.1016/j.obhdp.2005.09.002>
60. Podsakoff, P. M., & Todor, W. D. (1985). Relationships between leader reward and punishment behavior and group processes and productivity. *Journal of Management*, 11(1), 55–73. <https://doi.org/10.1177/014920638501100106>
61. Shaik, Z. (2011, June 22). Profane rant lands Southwest pilot in trouble. *Houston Chronicle*. <https://www.chron.com/news/houston-texas/article/Profane-rant-lands-Southwest-pilot-in-trouble-2082063.php>
62. Robinson, S. L., Wang, W., & Kiewitz, C. (2014). Coworkers behaving badly: The impact of coworker deviant behavior upon individual employees. *Annual Review of Organizational Psychology and Organizational Behavior*, 1(1), 123–143. <https://doi.org/10.1146/annurev-orgpsych-031413-091225>
63. Hu, Y., Strang, S., & Weber, B. (2015). Helping or punishing strangers: Neural correlates of altruistic decisions as third-party and of its relation to empathic concern. *Frontiers in Behavioral Neuroscience*, 9, Article 24. <https://doi.org/10.3389/fnbeh.2015.00024>
64. Leliveld, M. C., van Dijk, E., & van Beest, I. (2012). Punishing and compensating others at your own expense: The role of empathic concern on reactions to distributive injustice. *European Journal of Social Psychology*, 42(2), 135–140. <https://doi.org/10.1002/ejsp.872>
65. Adams, G. S., & Mullen, E. (2013). Increased voting for candidates who compensate victims rather than punish offenders. *Social Justice Research*, 26, 168–192. <https://doi.org/10.1007/s11211-013-0179-x>
66. Shnabel, N., & Nadler, A. (2008). A needs-based model of reconciliation: Satisfying the differential emotional needs of victim and perpetrator as a key to promoting reconciliation. *Journal of Personality and Social Psychology*, 94(1), 116–132. <https://doi.org/10.1037/0022-3514.94.1.116>
67. Bies, R. J. (1986). Interactional justice: Communication criteria of fairness. *Research on Negotiation in Organizations*, 1, 43–55.
68. Skarlicki, D. P., & Folger, R. (1997). Retaliation in the workplace: The roles of distributive, procedural, and interactional justice. *Journal of Applied Psychology*, 82(3), 434–443. <https://doi.org/10.1037/0021-9010.82.3.434>
69. Detert, J. R., & Burris, E. R. (2007). Leadership behavior and employee voice: Is the door really open? *Academy of Management Journal*, 50(4), 869–884. <https://doi.org/10.5465/amj.2007.26279183>
70. Cortina, L. M. (2008). Unseen injustice: Incivility as modern discrimination in organizations. *Academy of Management Review*, 33(1), 55–75. <https://doi.org/10.5465/amr.2008.27745097>
71. Cortina, L. M., & Magley, V. J. (2009). Patterns and profiles of response to incivility in the workplace. *Journal of Occupational Health Psychology*, 14, 272–288. <https://doi.org/10.1037/a0014934>
72. Liptak, A. (2017, October 2). Supreme Court divided on arbitration for workplace cases. *The New York Times*. <https://www.nytimes.com/2017/10/02/us/politics/supreme-court-workplace-arbitration.html>
73. Weale, S., & Batty, D. (2016, August 26). Sexual harassment of students by university staff hidden by non-disclosure agreements. *The Guardian*. <https://www.theguardian.com/education/2016/aug/26/sexual-harassment-of-students-by-university-staff-hidden-by-non-disclosure-agreements>
74. Patient, D. L., & Skarlicki, D. P. (2010). Increasing interpersonal and informational justice when communicating negative news: The role of the manager's empathic concern and moral development. *Journal of Management*, 36(2), 555–578. <https://doi.org/10.1177/0149206308328509>
75. Braithwaite, J. (2002). *Restorative justice & responsive regulation*. Oxford University Press.
76. Fehr, R., & Gelfand, M. J. (2010). When apologies work: How matching apology components to victims' self-construals facilitates forgiveness. *Organizational Behavior and Human Decision Processes*, 113(1), 37–50. <https://doi.org/10.1016/j.obhdp.2010.04.002>
77. Tomlinson, E. C., Dineen, B. R., & Lewicki, R. J. (2004). The road to reconciliation: Antecedents of victim willingness to reconcile following a broken promise. *Journal of Management*, 30(2), 165–187. <https://doi.org/10.1016/j.jm.2003.01.003>
78. Darley, J. M., & Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review*, 7(4), 324–336. https://doi.org/10.1207/S15327957PSPR0704_05
79. Mullen, E., & Okimoto, T. G. (2015). Compensatory justice. In R. Cropanzano & M. L. Ambrose (Eds.), *The Oxford handbook of justice in the workplace* (p. 477). Oxford University Press.
80. Malik, M., Wan, D., Ahmad, M. I., Naseem, M. A., & ur Rehman, R. (2015). The role of LMX in employees job motivation, satisfaction, empowerment, stress and turnover: Cross country analysis. *Journal of Applied Business Research*, 31(5), 1897–2000. <https://doi.org/10.19030/jabr.v31i5.9413>
81. Guchait, P., Kim, M. G., & Namasivayam, K. (2012). Error management at different organizational levels—Frontline, manager, and company. *International Journal of Hospitality Management*, 31(1), 12–22. <https://doi.org/10.1016/j.ijhm.2011.04.007>
82. Treviño, L. K., & Nelson, K. A. (1999). *Managing business ethics*. John Wiley & Sons.
83. Goodstein, J., & Butterfield, K. D. (2010). Extending the horizon of business ethics: Restorative justice and the aftermath of unethical behavior. *Business Ethics Quarterly*, 20(3), 453–480. <https://doi.org/10.5840/beq201020330>
84. Gromet, D., & Okimoto, T. G. (2014). Back into the fold: The influence of offender amends and victim forgiveness on peer reintegration. *Business Ethics Quarterly*, 24(3), 411–441. <https://doi.org/10.5840/beq20147814>
85. Brown, M. A. (2017). Being heard: How a listening culture supports the implementation of schoolwide restorative practices. *Restorative Justice*, 5(1), 53–69. <https://doi.org/10.1080/20504721.2017.1294792>
86. Boone, K. (2019, April 14). *LSU reinstates head coach Will Wade after meeting with school and NCAA officials*. CBS Sports. <https://www.cbssports.com/college-basketball/news/lsu-reinstates-head-coach-will-wade-after-meeting-with-school-and-ncaa-officials/>
87. Just, A. (2019, April 15). *LSU's Will Wade issues statement on reinstatement*. NOLA.com. https://www.nola.com/sports/article_702487ee-ac00-57cd-bc98-db3c188ff1a8.html
88. O'Reilly, J., Aquino, K., & Skarlicki, D. (2016). The lives of others: Third parties' responses to others' injustice. *Journal of Applied Psychology*, 101(2), 171–189. <https://doi.org/10.1037/apl0000040>
89. Skarlicki, D. P., & Kulik, C. T. (2004). Third-party reactions to employee (mis)treatment: A justice perspective. *Research in Organizational Behavior*, 26, 183–229. [https://doi.org/10.1016/S0191-3085\(04\)26005-1](https://doi.org/10.1016/S0191-3085(04)26005-1)

editorial policy

Behavioral Science & Policy (BSP) is an international, peer-reviewed publication of the Behavioral Science & Policy Association and Brookings Institution Press. 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 dual-review 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 ($\leq 2,500$ 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 ($\leq 3,000$ 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 ($\leq 4,000$ 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 ($\leq 5,000$ 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* ($\leq 5,000$ 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* ($\leq 1,000$ 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 Brookings Institution Press, 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.

sponsors

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



Alfred P. Sloan
FOUNDATION

BROOKINGS



MacArthur
Foundation

UCLAAnderson
SCHOOL of MANAGEMENT

THE WILLIAM AND FLORA
HEWLETT
FOUNDATION

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.

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 online 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

our mission

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.

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 the Brookings Institution Press.

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